FINAL REPORT

MORS/ITEA MINI-SYMPOSIUM

EMPHASIZING THE "E" IN T&E

1-3 OCTOBER 1991
NAVAL WAR COLLEGE
NEWPORT, RHODE ISLAND

General Chairman Mr Edward Brady

Technical Chairman Dr Patricia Sanders

Deputy Chairman Mr Phil Messing

This Military Operations Research Society mini-symposium proceedings faithfully summarizes the findings of a three-day meeting of experts, users, and parties intersted in the subject area. While it is not generally intended to be a comprehensive treatise on the subject, it does reflect the major concerns, insights, thoughts, and directions of the authors and discussants at the time of the mini-symposium.

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS NONE
2a. SECURITY CLASSIFICATION AUTHORITY n/a	3. DISTRIBUTION/AVAILABILITY OF REPORT Unlimited; Approved for Public Release.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE n/a	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Emphasizing the "E" in "T&E"	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Military Operations Research Society, Inc. 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION CNO, N81
6c. ADDRESS (City, State, and ZIP Code) 101 S. Whiting Street, Suite 202 Alexandria, VA 22304	7b. ADDRESS (City, State, and ZIP Code) The Pentagon Washington, DC 20350-2000
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office Naval Research 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-93-C-0061
8c. ADDRESS (City, State, and ZIP Code) Arlington, VA 22217	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO: O&MN; PROJECT NO: O&MN; TASK NO:; WORK UNIT ACCESSION NO:

11. TITLE (Include Security Classification)
Mini-Symposium Report -- Emphasizing the "E" in T&E

12. PERSONAL AUTHOR(S)
Patricia A. Sanders, Editor

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Proceedings	FROM 10/1 TO 10/3/91	October 1991	

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Test and Evaluation, Acquisition Process, Analysis Tools, Suitability Testing

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This is the Proceedings of the joint Military Operations Research Society (MORS) and International Test and Evaluation Association (ITEA) mini-symposium on emphasizing the "E" in T&E. The mini-symposium was held at tne Naval War College, Newport, RI, on 1, 2, and 3 October 1991. In addition to a summary of the mini-symposium's findings, individual sections summarize discussions and findings from the six working groups: Evaluation as a Tool for Test Planning and Improved Test Execution; Evaluation Framework to Close the Loop in the Acquisition Process; Evaluation during Development Testing to Reduce Risks; Evaluation Techniques to Overcome the Limitations in Suitability Testing; Evaluation as a Critical Element of the Test and Evaluation of Evolutionary Acquisitions; Using Testing to Enhance the Credibility of Our Analysis Tools.

20. DISTRIBUTION AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Pierce J. Johnson	22b. TELEPHONE (Include Area Code) 703-697-8051 22c. OFFICE SYMBOL N811, OCNO

DD Form 1473, JUN 86 Previous editions are obsolete SECURITY CLASSIFICATION OF THIS PAGE

S/N 0102-LF-014-6603

The **Military Operations Research Society (MORS)** is the professional association of military operations research analysts and users of military operations research from both the military and civilian sectors. MORS is sponsored by:

- The Deputy Under Secretary of the Army (Operations Research)
- The Director of Program Resource Appraisal, Office of the Chief of Naval Operations
- The Director, Directorate of Programs and Evaluation, Headquarters US Air Force
- The Director of Force Structure, Resource and Assessment, The Joint Staff
- The Assistant Secretary of Defense (Program Analysis and Evaluation)

Under the Contractual Sponsorship of the Office of the Chief of Naval Research

The **International Test and Evaluation Association (ITEA)** is a not-for-profit professional organization dedicated to furthering the professional and technical interests of the test and evaluation community.

PLANNING AND ORGANIZING COMMITTEE

General Chair:	Mr. Ed Brady Strategic Perspectives, Inc.
Technical Chair:	Dr. Patricia Sanders Office of the Assistant Secretary of Defense (Program Analysis and Evaluation)
Deputy Chair:	Mr. Phil Messing The MITRE Corporation
ITEA Advisor:	Mr. John Bolino Office of the Deputy Director Defense Research and Engineering (Test and Evaluation)
MORS Advisor:	Dr. Marion Bryson US Army Operational Test and Evaluation Command
Site Coordinator:	Mr. John Keegan Naval Underwater Systems Command
Administrative Coordinators:	Mr. Dick Wiles/Ms. Natalie Addison Military Operations Research Society
	Mr. Alan Plishker International Test and Evaluation Association

TABLE OF CONTENTS

OVERVIEW AND SYNTHESIS

EMPHASIZING THE "E" IN T&E

BACKGROUND

The phases of developmental (or technical) and operational testing are critical elements of assessing the performance and effectiveness of weapons systems in the acquisition process. As such they receive considerable attention. The same emphasis has not generally been placed on evaluation. The Defense Science Board (DSB) 1989 Summer Study Task Force on Improving Test and Evaluation Effectiveness observed that:

> "...the test and evaluation community places a heavy emphasis on test and a light emphasis on evaluation. Test and evaluation are interrelated and complementary processes, both of which are necessary; neither alone is sufficient. Evaluation must be used to judge overall system performance against the operational mission requirements and to reassess performance as the mission requirements and system design evolve. This evaluation is supplemented by test results."

This Task Force and numerous test and evaluation documents have pointed out the need for a test and evaluation process that uses an evaluation framework established at the start of the system acquisition program. The evaluation framework would establish probable evaluation procedures and, as the program progresses, the framework would be upgraded consistent with the advancing state of knowledge concerning the system or evaluation methods.

Historically, the focus of evaluation has been on evaluating the results of testing against the given requirements or criteria. As a result, the evaluation planning, to include selection of evaluation methodology, is driven by or follows test planning instead of evaluation being used to pinpoint the critical aspects of the system that should be tested and under what conditions testing ought to be conducted.

Evaluation planning should be required prior to testing to identify the most significant variables to be measured during the various testing phases. After testing, evaluations are required to examine the data in detail, as well as support examinations of the extremes of the test envelope. System performance during operational situations not replicable on the range can often be inferred from these examinations.

Placing emphasis now on evaluation is particularly necessary and timely because of the increasing sophistication of weapons systems, fiscal constraints with the corresponding requirement to optimize the effectiveness of test and evaluation, and the currently ongoing reassessment of the acquisition process including the management and resources for test and evaluation. There needs to be a big "E" outlook as well as the little "e" which is the natural planning and assessment process that

has always been an integral part of test and evaluation.

OBJECTIVE

The objective of this mini-symposium was to provide a forum in which the analytic community and the test and evaluation community could come together to explore the philosophy, policy, processes, and methods of evaluation. The forum:

- Discussed the need for more thorough evaluation of systems in the defense acquisition process in the current and evolving environment;
- Examined the role of evaluation in the total test and evaluation process;
- Explored the role that the analyst plays in the acquisition, particularly test and evaluation, process;
- Explored the role that the tester plays in the evaluation process throughout the acquisition cycle;
- Examined and proposed evaluation techniques and tools.

APPROACH

The mini-symposium was challenged at the outset by the keynote speaker, Mr. Walt Hollis, Deputy Under Secretary of the Army (Operations Research), and a panel of senior decisionmakers and members of the test and evaluation community who provided their perspectives on the subject of the meeting. The participants then approached the above objectives principally through six parallel working groups where individual evaluation topics were examined in depth. These sessions included informal presentations, discussions, formulation of issues and approaches to resolving these issues. The areas of focus of these working groups were the following:

I. **Evaluation as a tool for test planning and improved test execution.** Given the complexity of systems being tested and the environments in which they need to be tested, how can we use pre-test analysis to better plan for the most efficient use of resources and the most insightful test results.

II. **Evaluation framework to close the loop in the acquisition process.** This was intended to address a possible architecture to link mission area analysis for requirements definition, cost and operational effectiveness analysis, and the analysis of the results of test and evaluation.

III. **Evaluation during development testing to reduce risks.** How can improved evaluation be used to lessen the risks inherent in the development of complex weapons systems and to prevent unfortunate surprises when the system undergoes operational testing.

IV. **Evaluation techniques to overcome the limitations in suitability testing.** Using evaluation to improve our ability to predict the reliability, maintainability and availability of weapons systems and reduce life cycle costs.

V. **Evaluation as a critical element of the test and evaluation of evolutionary acquisitions (EAs).** For the types of systems where EA is the preferred acquisition

strategy—command and control systems, software intensive systems,...—how can we better assess effectiveness through the use of evaluation techniques.

VI. **Using testing to enhance the credibility of our analysis tools.** How can we better use the results of tests and exercises to "validate" or raise the level of confidence we place in the tools we use for evaluation.

Each group developed conclusions and recommendations on its specific topics and findings relevant to the overall conference theme which were presented in a closing plenary session of the mini-symposium. This enabled all participants to share the insights gained by each group and prevented any group from maintaining too limited a perspective about its focus topic. Summaries of the working group deliberations are included in this report. This material is a synopsis of the discussions as seen by the working group co-chairs and does not purport to be a verbatim representation of the views of any one of the participants.

While each group had its own unique focus, the topics were not mutually exclusive. There was an intersection of the issues that were addressed in each working group, as well as a number of common and related results. In fact, several recurring themes were evident. The principal observations, findings, and recommendations are provided below, but many thoughtful and more specific insights and suggestions for improving the process are found in the individual working group reports.

OBSERVATIONS/FINDINGS

1. **The integrity or "oneness" in the acquisition process must be significantly improved for that process to be effective.**

The acquisition process is complicated, extended over time, and has many players. The evaluations that occur must form a whole, with all the pieces consistent. A procedure is missing to ensure a logical relationship between the Mission Area Analysis (MAA), the Mission Need Statement (MNS), the Operational Requirements Document (ORD), the Cost and Operational Effectiveness Analysis (COEA), and the Test and Evaluation Master Plan (TEMP). Lack of connective tissue, links, logic, and rationale in the hierarchy of evaluations severely handicaps the decisionmaker. On the other hand, if these analyses form a coherent set, each evaluation will be stronger than it is today and the acquisition process will be more productive.

2. **Lack of communication is a major problem in the evaluation process.**

A primary cause of the lack of traceability throughout the acquisition process is the diversity of organizations responsible for the different evaluation products. A logical relationship between the MNS, ORD, COEA, development test and evaluation and operational test and evaluation, is difficult—if not impossible—to achieve unless the organizations responsible for developing each have a better working relationship and communicate early in the process. Better coordination among evaluators and users would increase the chance of getting the statement of precise testable requirements necessary for effective test and evaluation.

3. **The Test and Evaluation Master Plan (TEMP) is deficient as a planning document and management tool for test and evaluation.**

While the TEMP identifies test activities and resources, its orientation is such that it addresses evaluation poorly, if at all. It does not include the rationale for each test activity nor the information that is to be gained. Its format specifically encourages a separation of development and operational test activities. Test resources are identified while evaluation resources are frequently ignored. The logical relationship between T&E and the MNS, ORD, COEA, and evaluation measures of effectiveness (MOEs) is not included.

4. **We frequently neither conduct the right tests nor collect the right data.**

Scarce resources are wasted conducting tests to collect data that are never used in evaluation. Questions that need to be addressed are not well defined and test objectives are unclear. Decisionmakers need information, not data. The lack of the necessary information for confident decisionmaking produces a negative impact on the acquisition process, disrupting program schedules, modifying funding streams, requiring additional program reviews, and causing other severe disruptions to the process.

5. **Most test and evaluation programs are too narrowly focused.**

Conclusions are often based on the results of a single test or series of tests rather than being broad based on all relevant test data and appropriate support from models and simulations. Many sources of information outside the formal test arena are not considered in any comprehensive evaluation of the system. Even within the T&E community, neither development tests nor operational tests maintain the proper cognizance of the other's efforts and many opportunities for shared information and enhanced insights are lost. Evaluation is neither begun early enough in the process, nor continued after formal operational testing is complete.

6. **System deficiencies are too often identified later in the test and evaluation process than is necessary.**

A large number of deficiencies uncovered in operational test could have been discovered much earlier in the development cycle with better and less expensive opportunities for early correction. Waiting until late in the acquisition process to identify problems is costly. One of the greatest challenges to evaluators is in assessing the risk arising from uncertainty as to whether early designs or architectures with limited functionality can eventually support the full functionality as established in the system's capability objectives.

7. **Evaluation capabilities are not well resourced.**

The lack of perceived early and coherent planning for evaluation contributes to a corresponding lack of timely identification of evaluation resources. Earlier definition of required analytical tools, including modeling and simulation, is essential if they are to be available when needed to support the process. Validation, verification, and accreditation of these tools does not occur, therefore the results of their use is often not perceived to be creditable.

Testing, particularly on the development side of the community, has become a mature science, and significant investment has been made in facilities, testing hardware, and instrumentation. But existing test resources frequently do not satisfy the data collection needs of the evaluators. And there is a growing recognition that the resources for trained people to do the analysis are becoming scarce.

8. **There is a need to increase management attention and create better structures for the conduct of suitability evaluations.**

A review of 26 systems, conducted at the request of the Under Secretary of Defense for Acquisition in October 1990, confirmed that suitability concerns are much more prevalent than effectiveness concerns at the time of the full-rate production decision. The leverage for identifying and fixing suitability problems early has been documented at well over a 10-to-1 return on the investment ratio. Yet suitability evaluation is under-utilized, particularly in developmental testing.

9. **Existing DoD and Service regulations do not provide adequate guidance with respect to evolutionary acquisition (EA) and, under some interpretations, inhibit or even preclude EA as a strategy.**

EA is a strategy which has the potential for permitting the responsible discharge of duties toward articulating system requirements and conducting adequate test and evaluation without postponing the joy of early deployment and incremental procurement of useful capabilities. EA can actually become a risk reduction strategy in that a commitment is made sequentially to relatively small increments of achievement, rather than gambling on the ability to accomplish a single ultimate goal. Challenges to its successful application come from uncertainty in the development community as to its utility, mistrust in the oversight community of its "legality", and discomfort in the T&E community regarding how to carry out their responsibilities.

RECOMMENDATIONS

1. **A procedure should be established to better link the MAA, MNS, ORD, COEA, and TEMP.**

Prior to Milestone 0, an evaluation oversight group should be established, populated by those responsible for the generation of requirements, the cost and operational effectiveness analysis, and the technical and operational test and evaluations, to oversee the life cycle evaluation of weapons systems. The group should assure that methodologies, rationales, and evaluations are consistent.

A better thought out requirements process should be established that includes cost-benefit and trade-off analysis from the COEA/MAA. The MNS should be strengthened by better analysis at the start of the process. The COEA should be complete in the sense of including sensitivity analyses and all relevant, affordable options. The TEMP should include or reference the COEA measures of effectiveness and suitability at each TEMP update. Both should identify and provide rationale for changes in scenario/threats and related requirements. All should be logically traceable from the MAA.

All documentation, in approved form, should be available at the milestone decision points.

2. **There should be a mechanism for early and continuous coordination between the Services, OSD evaluators, PA&E, DOT&E, DDDR&E(T&E), and the JROC.**

This could take the form of a Test and Evaluation Integration Working Group; it could be an extension of a COEA Study Advisory Group; or it could be an Evaluation Oversight Group as recommended above. Better communications are preferred to more bureaucracy: the important thing is to get the people together. Use of the Test and Evaluation Community Network (TECNET) should be promoted. Interservice agreements from Project Reliance should be capitalized upon to open communications among service evaluators.

This coordination must take place before Milestone I, and preferably before Milestone 0. Measures of performance and effectiveness should be fully coordinated with the user. The early stages are exactly where everything can be set on the right track or left to wander.

There should be a better demonstration by OSD of a willingness to work with the Services early in the process to help get the job done.

3. **A single evaluation framework is needed. Serious consideration should be given to replacing the TEMP with a document which provides a single, integrated EVALUATION Plan.**

The Master Evaluation Plan (MEP) should be considered as the capstone in which all evaluation requirements are coordinated and from which all activity in support of the evaluation is derived. This document should be the joint responsibility of the evaluation community—not the Program Manager. It should lay out the evaluation framework prior to any selection of tests and support with matrices, data needed by evaluators. Evaluation resources should be identified.

The MEP should clearly emphasize the plan for evaluation as the first and foremost requirement in the test planning and execution process. Second, it should shift the emphasis from purely test oriented activities to meeting information needs with more comprehensive evaluations incorporating all sources of data. And third, it should provide a clear road map for test requirements based on the needs of the evaluation.

4. **An iterative looping of Evaluate-Test-Evaluate cycles should be used to continuously and comprehensively plan, execute, and report the performance and effectiveness of a system.**

The Evaluate-Test-Evaluate process, as described in the report of Working Group I, should provide a continuous and comprehensive assessment of the performance, effectiveness, and suitability of a system. Feedback loops should not only provide timely information to decisionmakers, but also help to ensure that test planning and execution meet the evaluators needs, that the correct data is collected, and that neither too much nor too little testing is accomplished.

Rehearsal of data collection should be an important aspect of this process.

Availability of test and evaluation resources should be identified early and report formats should be established prior to test. Pilot tests should be encouraged to ensure that adequate experimental design supports the process.

Tools and techniques should be developed to support this process.

5. **Evaluation should be undertaken as a continuous and comprehensive process.**

Evaluation should be recognized as a continuous process that extends from earliest concept to post initial fielding—even after formal operational test and evaluation is completed. It should be issue and decision oriented rather than calendar oriented. Evaluation should continue to determine the system's viability in light of new and changing threats, new doctrine and tactics, technology breakthroughs, or for possible new applications or missions.

Policies and resources should be provided for continuous evaluation including a requirement for periodic reports. Built-in-test equipment and data recording should be incorporated into the design of military systems whenever appropriate. Databases should be archived and made available to evaluators throughout the system's life. A data management system should be created and storage facilities should be provided.

Evaluation should intelligently combine data from all sources: formal testing, mock-up examinations, field experiments, work cycle examinations, experiments with surrogates, attendance at design reviews, modeling and simulation, etc. Greater sharing of developmental and operational test data should be facilitated, rather than obstructed as it is now.

6. **Evaluation techniques should be used early to identify both risks and opportunities.**

Early evaluation concepts and techniques should include provisions for evaluating components as a part of the larger system, e.g., system compatibility, software interoperability and protocols, increases in operational burden, changes to electro/optical signatures, etc.

Experiments with surrogates should be used to lead to an early understanding of whether or not a new technical opportunity has operational utility. Advanced simulation technologies such as virtual reality and the "electronic battlefield" should be developed and exploited.

7. **Evaluation capabilities should be adequately resourced.**

Required evaluation capabilities should be identified early in the program, resourced, and developed throughout the life cycle of the program.

A clear understanding should be developed of the use of modeling and simulation for use in T&E. A library of appropriate models and simulations should be established; interoperability of models and simulations should be encouraged; standardized access to data bases should be developed; and facilities should be provided for running simulations. System models should be obtained as early contract deliverables.

Credibility of evaluation tools should be given greatly increased attention.

Verification, validation, and accreditation (VV&A) should be required for models and simulations used in evaluation of weapons systems. The Defense Modeling and Simulation Office should coordinate activities to develop and promote VV&A techniques and activities. The results of tests and exercises should be used to raise the confidence we place in our evaluation tools.

Investments in new test capabilities should be tied to evaluation requirements. Investments should be driven by technologies vice specific programs, with technology edge and common requirements funded by the Office of the Secretary of Defense.

An investment should be made in education and training for evaluators, to include specialized training in operations research, systems analysis, systems engineering, risks manufacturing processes, and operational exposure. An evaluator's syllabus should be developed for the Defense Test and Evaluation Professional Institute (DTEPI) and consideration should be given to setting up a test and evaluation degree program.

8. **Management attention to suitability evaluation must be increased and sufficient structures must be in place to do so.**

An OSD focal point for suitability evaluation should be identified, perhaps by creating positions within USD(A&T)/D,T&E and DOT&E, to take the lead in planning and overseeing the execution of the process. Guidance should be developed and provided with respect to:

- Standardized definitions, data collection, and data base design, particularly associated with RAM;
- Application of reliability growth, e.g., update of MIL HDBK 189, Reliability Growth Management;
- Reporting of RAM factors and measures, e.g., point estimates versus confidence intervals;
- Evaluating suitability where contractor support is planned.

The impact of suitability evaluations should be increased through such actions as linking to cost and operational effectiveness analyses, better use of modeling and simulation, and improving the way we measure the elements of operational availability.

Greater attention should be paid to evaluating software RAM, including a significant and immediate acceleration in DoD-wide emphasis, coordination, and incorporation of evaluation methodologies and metrics for evaluating software.

9. **DoD and Service guidance should be provided with respect to evolutionary acquisition.**

Crisp criteria should be provided for deciding whether a particular system is appropriate for EA and for early identification of the essential, militarily useful core capability of a system that will serve as the nucleus for further evaluation.

An appropriate test and exercise environment should be created, to include a prototype of the evolving system, simulators,

stimulators, and replicas of interoperating (or adversarial) systems, instrumentation, users or their surrogates.

Measures of performance and effectiveness should be defined to grade system progress toward full capability, while distinguishing between threshold and mature capabilities. A disciplined, visible mechanism for providing evaluation feedback should be established and exercised, both to effect the design of the current system increment and for the requirements refinement process.

Recommendations One through Eight above should be reiterated for systems with an EA strategy—at each step in their evolution.

SUMMARY

The participants in the mini-symposium consistently presented evidence for increased emphasis on evaluation with a big "E" in the test and evaluation process—to decrease both the risks and costs in a very complex decision process. A number of general and specific recommendations were developed for improving our evaluation process. All of them are achievable, although not necessarily within the test and evaluation community itself. They should be viewed as an opportunity which, if embraced, could realize substantial benefits for the Department of Defense.

EXCERPTS FROM THE KEYNOTE ADDRESS TO THE

MORS/ITEA MINI-SYMPOSIUM

EMPHASIZING THE "E" IN T&E

Mr. Walter Hollis
Deputy Under Secretary of the Army (Operations Research)

I came to have a close direct involvement in testing in 1968 when I became the Scientific Advisor to the Commanding General of the Combat Developments Experimentation Command at Ft. Ord, California. Previous to this assignment I had been a materiel developer for a long period of time. I suppose my point of view about testing was similar to that of other materiel developers at that time. Testing was akin to a final exam in college but different in the sense that, in those days, there had been no quizzes along the way. My experience at CDEC opened my eyes to the possibilities for a greater interaction between the materiel developer, and the ultimate user of his equipment with the tester as the facilitator for the interaction. It also became clear to me while at CDEC that those who execute tests, while having the responsibility to report that "which happened on the range," the conditions under which the data were gathered, and any facts which might impact upon the utility of that data, were not themselves in the best position to generalize from that data in an evaluative sense. In fact, one of the instructions I had been given by the command group at CDC when I went to CDEC was to take action to make the reports of experiments clear as to what data was taken from the field and what data had been generated by computers.

When I took up my assignment at OTEA, the Agency was in its formative stages. The fact of, or the existence of an OTEA was not generally accepted. I found the preliminary organizational arrangement at OTEA to be parallel to that of the Army branches, i.e., artillery, armor, etc. These organizational entities contained test designers, test executors, and evaluators. There was, however, no concrete idea as to what an evaluation should be. One of the first actions Ray Ochs, the Commanding General, and I took was to realign the organization into Test Design, Field Test, and Evaluation Divisions, with a Technical Support Division to support the entire agency. Among other objectives for this realignment we wished to create an internal adversarial situation between the designers, executors, and evaluators. OTEA needed to establish its credibility quickly. We felt these arrangements would help us to do so.

In keeping with our need to establish our "bonafides," we initially set about to do evaluations based almost entirely on the data we had gathered in the operational tests. Although admittedly a narrow focus, this approach to evaluation, which I now term the "little e," supported our precepts for the agency of adequacy, quality, and credibility.

As time passed, it became evident that the narrow "little e" approach to evaluation would not remain viable. There were systems under development for which credible operational tests, which were affordable, would not be possible. Among these were systems such as the Maneuver Control System designed to assist corps, division, and brigade staffs in the execution of the command function. There were systems where achievement of useful estimates of both effectiveness and reliability could not be achieved in a single or in some cases multiple tests. Such systems included the large missiles such as the PATRIOT and non-nuclear Lance. Clearly, simulation would have to be employed for adequate effectiveness evaluations and pooled data over all tests would be required for good reliability estimates.

When I assumed my present duties and became a member of the Army System Acquisition Review Council, I also realized that the OTEA evaluations were not always helpful to that body in that findings of a failure of a system to meet this or that criterion or failure to meet one out of several criteria were not accompanied by an answer to the "so what" question?

Digging further, I found that Army senior leaders saw that:

1) The major portion of the findings from operational testing relating to soldier system interaction could have been uncovered much earlier in the development cycle with corresponding opportunities for early corrective action.

2) Estimates of expected combat availability, operational reliability, maintenance burdens, etc., were done too early in the acquisition cycle.

3) Evaluations were too narrowly focused on the results of a single test rather than being broad based on all relevant test data and appropriate support from models and simulations.

Based on those findings and on my own concerns which were quite similar, I proposed to the Army leadership that we change our operational test agency's evaluation philosophy. That is, that the "E" in evaluation be written large rather than small. To implement this new philosophy it was necessary to:

1) Recognize that evaluation should be a continuous process extending from earliest concept to post initial fielding.

2) Recognize that evaluation and "testing" should be issue rather than calendar-oriented. This is to say that evaluators should take all opportunities whenever these arise to cause issue-oriented data to be generated as opposed to waiting for the arrival of a calendared OT period. This is not to say that a formal IOT&E is not required. Indeed it is and must be a part of the process.

3) Recognize that much useful data relating to human interface issues and some early insight into supportability issues can be derived from examination of work cycles, experiments with surrogates, and attendance at design reviews.

4) Recognize the need for a post-fielding examination of broad supportability issues since, in considerable measure, these issues turn on the ability of the next higher

level system to support a fixed number of new item level systems together with whatever other item level systems must also be supported by next higher level systems.

5) Recognize the need for our decision criteria to be appropriate to the level of maturity of the item level system under evaluation, i.e., we should not necessarily expect the performance or the reliability of a system in advanced development to have reached the levels we expect to achieve at maturity. (Incidentally, our specifications should reflect this also.)

6) Recognize that field experiments with surrogates can lead to a very early understanding of whether or not a new technical opportunity has operational utility and, if so, how it might best be employed, supported, etc.

Underlying all of the above is the concept that relevant data come from many places. Some come from what we understand as the "classical OT;" some come from mock-up examination, field experiments, and sample data collection systems, post fielding. What is required is that a relevant process generates the data, the data are well enough recorded so that the process which generates the data may be reconstructed from the data, and that a proper methodology is developed to utilize all available data in a proper context.

We implemented this philosophy in late 1983. It is called Continuous Comprehensive Evaluation (C2E). Many of the Army systems deployed and employed in OPERATION DESERT SHIELD and OPERATION DESERT STORM have been evaluated by this process. It is, of course, a matter of conjecture as to the degree this process contributed to the outstanding performance of the equipment. Personally I believe that it did have a positive impact.

PANEL PRESENTATION

CHALLENGE TO THE PARTICIPANTS: WHY EMPHASIZE THE "E" IN T&E?

PANEL MEMBERS:

OSD Decisionmaker's Perspective:	**Mr. Richard Ledesma** Acting Deputy Director Defense Research and Engineering for Test and Evaluation
Program Manager's Perspective:	**Mr. John Kelley** PATRIOT Program Office Raytheon
Operational User's Perspective:	**Major General James Drummond** Consultant (former Commanding General, US Army Operational Test and Evaluation Agency)
Tester's Perspective:	**Dr. Marion Williams** Chief Scientist Air Force Operational Test and Evaluation Center
Planner's Perspective:	**Dr. Philip Dickinson** Director Future Tactical Intelligence Systems E-Systems

SUMMARY

With more than twenty years experience in all aspects of test and evaluation in the Office of the Secretary of Defense, **Mr. Richard Ledesma** emphasized the point that testing produces data and that decisionmakers do not need data, they need information. Evaluation of the data provides the information upon which decisions must be based. The decisionmaker does not have the time to sort through raw data seeking the information he needs. He should have analysis and recommendations from his staff

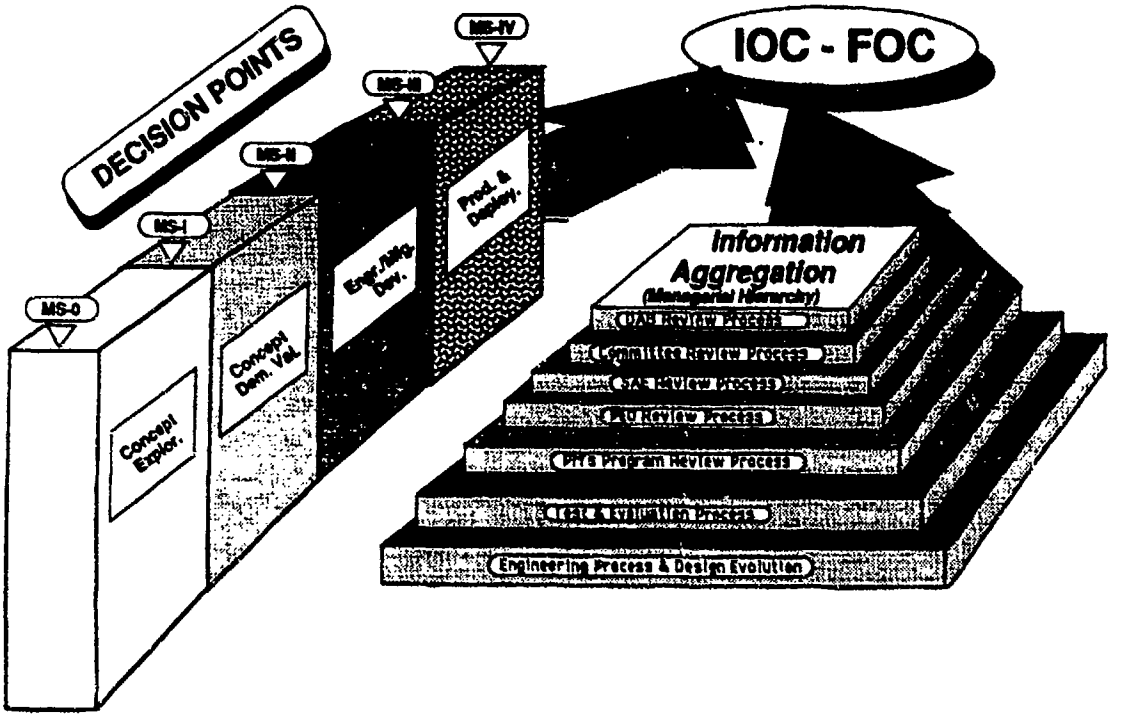


Figure 1. Changing Information Needs

of analysts. He needs to know risks, program maturity, what the requirements are versus actual performance, and the impacts thereof, etc.

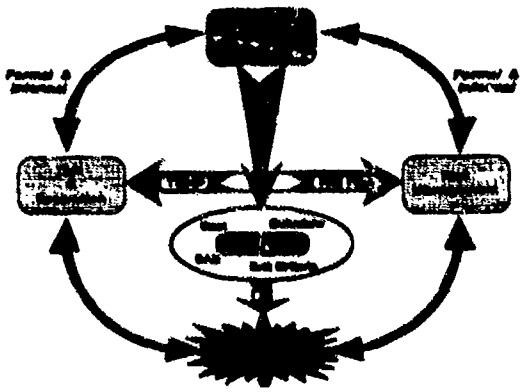
He pointed out that at each management level and at each decision milestone, information needs change in emphasis and complexity. (Figure 1) The focus is on the goal of fielding the system on time, under cost, and meeting the user's requirements. The decisionmaker must confront formally at each milestone, the aggregate of information needed by the full hierarchy of management on all aspects of the program for the decisionmaking process. One of the constants in the information needs of decisionmakers throughout the acquisition process and at all levels is the desire for accurate and timely results from test and evaluation, including the status of system risk.

Mr. Ledesma specifically touched on the topics of each of the working groups. With respect to the needs for evaluation to support the test planning process, he said that the key is, given the complexity of the systems being tested and the environments in which they need to be tested, to better plan for the most efficient use of resources and the most insightful test results. Such evaluation needs to consider "testability." Both the weapons system and the test itself should be evaluated to ensure positive contribution to the overall process. His bottom line was that if T&E results are not providing decisionmakers with the required

information for confident decisionmaking, then management will impact the acquisition strategy by modifying funding streams and program schedule, requiring additional program reviews, establishing or modifying exit criteria, and taking other actions that cause severe disruptions to the development process.

He addressed risk management and stated that the T&E strategy and the risk management strategy must be in harmony and support each other to provide the decisionmaker the critical information they need to make informed decisions.

Figure 2. T&E and Risk Management



With respect to suitability, **Mr. Ledesma** was critical of both the way in which suitability was designed into a system and the way developmental testing approached it. He singled out an attitude of "waiting until operational testing discovers problems" as being too costly and not supporting the acquisition process. **Mr. Ledesma** cited a review of how DoD ensures it is building suitable systems for our fighting forces requested in October 1990 by Mr. Yockey, Under Secretary of Defense for Acquisition. An examination of 26 DOT&E Beyond-Low Rate Initial Production (B-LRIP) Reports confirmed that suitability concerns are more prevalent than effectiveness concerns at the time of the full-rate production decision. His challenge was to improve our evaluation techniques to overcome limitations in suitability testing, improve our ability to predict reliability, availability, and maintainability of weapons systems, and, therefore, reduce life cycle costs.

Mr. Ledesma also spoke to the improving of evaluation tools with two approaches:

- Using test results and exercises to improve our analytical tools and raise the level of confidence in their credibility.

- Improving the methodologies and technologies inherent in our evaluation tools to also raise confidence and credibility levels.

He covered the discussions of modeling and simulation at a recent T&E Offsite Meeting focusing on these issues:

- Development of a clear mission statement of modeling and simulation in test and evaluation.

- Development of a methodology for determining the bounds of modeling and simulation for use in T&E.

- Identification of a need for interoperability among (1) T&E models and (2) data bases and standards and protocols to support such interoperability.

- Recognition of the conditions under which developers, development testers, and operational testers can use common models and simulations in test and evaluation.

- Encouragement of earlier definition of modeling and simulation needs for test and evaluation.

Mr. Ledesma closed by expressing his belief that while many of the problems in test and evaluation were beyond the analysts control, much improvement was possible by emphasizing the "E" in T&E.

Mr. John Kelley, from his perspective as a manager within the PATRIOT Program Office since 1984—including managing the PATRIOT systems test program—presented a number of challenging concepts for enhancing evaluation in the T&E process. The first of these was to employ a "transition to testing" concept similar to what is done in transitioning to production. To accomplish this he felt that there needed to be the involvement of the designer, the analyst, the evaluator, the test planner, and the test conductor from day one. By intentionally planning for transition to test, a program avoided the "throw it over the wall" attitude. **Mr. Kelley** also advocated a blending of analysis/evaluation organizations with testing organizations.

Another insight **Mr. Kelley** provided was to encourage the use of validated simulations to minimize costly testing at the outer edge of a systems's capability. He provided an example from the PATRIOT program involving extensive testing at a hardware-in-the-loop simulation facility to compensate for limited flight test at performance boundaries, particularly in the electronic countermeasures arena. His example illustrated a cost-effective approach to the use of such validated simulations.

Mr. Kelley also emphasized the need to use evaluation to ensure that one was conducting the right test. His maxim: If a test will result in data which when evaluated will not answer the mail—reconsider! He provided an example related to a live fire exercise from the PATRIOT FOE III to illustrate his point. Related to this he discussed the need to also ensure that one collected the correct data. He stated that every test was a potential source of data; he emphasized the need for evaluation after formal operational testing is concluded; and he encouraged testers to collect data for evaluation even in purely tactical situations.

Major General James Drummond's Army career included more than twenty years in the materiel acquisition business and included serving as the Commanding General of the TRADOC Combined Arms Test Activity and as the Commanding General of the US Army Operational Test and Evaluation Agency (OTEA). He began his remarks by stating that he has long been convinced that the T&E community has been too focused on the "T"—the testing of systems, to the neglect of the "E"—the evaluation.

General Drummond said there have been obvious and understandable reasons for this. Testing, particularly on the development side of the community, has become a mature science. We have made significant investment in facilities, in testing hardware and in instrumentation to measure precisely a wide range of physical properties and operating characteristics. We have developed a family of specific developmental

tests that measure and record everything from ability to withstand shock and vibration of EMP and a variety of other external phenomena, to operations in extreme environmental conditions.

During operational tests, **General Drummond** contended, the first real opportunity to observe effectiveness and suitability when in the hands of troops, everyone connected with the acquisition has his eye not only on how the system is performing, but also on how objectively and fairly the test is being conducted. So with this service-wide interest and visibility into the test, their high cost and, if not a direct impact on readiness, at least an inconvenience to troop units, it is only human nature that we emphasize the test and lose sight of the fact that evaluation is the end game! Even formal milestone tests, as important as they are, are simply events along the way which provide additional data and insights to the overall evaluation of system effectiveness and suitability.

General Drummond also noted that there are many other sources of information outside of tests and, in fact, much information may never appear in the results from any test. Yet this information must be considered in any comprehensive evaluation of a system. This is one reason he so strongly supports the concept of Continuous Comprehensive Evaluation (C2E) which the Army institutionalized some five or six years ago. Evaluation, he considered, is an art and perhaps it will never be a science; nevertheless C2E implies a scientific methodology is applied to all of these outside-of-test data and information. The informed comprehensive evaluation that all this provides is what the decisionmakers are after when they reach decision points.

Additionally, **General Drummond** stressed organizational steps that could be taken and expressed satisfaction with the recent reorganization of the Army's Operational T&E world giving evaluation its proper place and the focused assets for the job.

General Drummond offered for the mini-symposium's consideration what he called the Master Evaluation Plan. The Test and Evaluation Master Plan (TEMP) is supposed to be "the comprehensive road map of all the testing which a system sill undergo during its development cycle." He had three criticisms of the TEMP:

- First, the developer of the TEMP is the Project Manager who has a natural interest in minimizing the exposure of the developmental system to those who would challenge it. For some PMs, their test strategy seemed to be test avoidance. Therefore the "Keeper of the TEMP" was the absolute wrong guy!

- Second, for literally hundreds of non-major systems, what we often call "horse shoe nail" systems, there is no TEMP at all. The immediate problem is that no one can program assets for test and evaluation.

- And last, while the TEMP is a road map of where you intend to go in testing, it doesn't tell you why you are going through each test or what information you expect to gain at each stop. Further, this road map doesn't convincingly lay out test "travel plans" to pick the most economic or scenic route from the standpoint of issues critical to and/or data requirements common to both

the development and operational evaluation communities.

General Drummond provided the hypothesis that the basic document to guide the entire T&E process should be a Master Evaluation Plan (MEP), the development of which would be the joint responsibility of the independent operational and development evaluators. This document would lay out the evaluation framework prior to any selection of tests and support with matrices, data needed by both evaluators. A MEP would introduce consideration of operational issues into earlier testing and could lead to much earlier identification of problem areas. More importantly, it would get evaluation planning out in front of test planning where it properly belongs.

As the Chief Scientist of the Air Force Operational Test and Evaluation Center where he has been a major influence on the T&E community since 1974, **Dr. Marion Williams** had a number of additional issues to raise for the consideration of the participants. First of all, he brought up the need for a specific, well-defined question to answer as an important factor for effective testing. Too often, he felt, the necessary evaluation of the military requirement that the weapons system was being developed to meet had not been adequately accomplished. Designing, executing, and reporting on a test whose objective is not clearly understood is a futile exercise.

In a similar vein, **Dr. Williams** challenged testers to perform better evaluation for test planning, stating that the test was not the end product, but merely provided input to the evaluation. He cited what he called "brute force testing" as an easy way out. In this approach, the test is accomplished, ALL available data are gathered, and the evaluator hopes the conclusions fall out. Efficient test planning, he contended, was much different. It requires up front thinking and the development of an evaluation architecture with a simple, logical structure. T&E is scoped to address the specific questions at hand. The risks are identified and testing is accomplished in the risk areas. Analytic tools are used to increase testing efficiency. Simulation can be used to identify performance sensitivities and areas of uncertainty. **Dr. Williams** felt that such an approach was often discussed, but not often accomplished.

Dr. Williams next raised some issues related to the evaluation of test data. He made the point that test and evaluation are of no value if the results are not effectively communicated. The result of a test should not be a report or a briefing, but the communication of information that is relevant to a decision. He also stressed the need to get as much information as possible out of the available data. We must be able to combine intelligently the data coming from multiple sources: development test, simulation, operational test. We must know how much testing is enough. And we must develop the analytical tools to allow us to do all this.

Some additional challenges were issued by the final member of the panel, **Dr. Philip Dickinson**. Drawing on his previous experience as the Technical Director of US Army OTEA and Deputy Assistant Secretary of the Army for Requirements and Planning, as well as his current role in industry, **Dr. Dickinson** talked about the need to evaluate the process. He stated that this was perhaps as important as the evaluation of the system

under development. He presented several case studies illustrating deficiencies in requirements, operational concepts, the role of a system within the force, training and resourcing. He encouraged the participants to use evaluation to challenge the requirements. He cited examples including the Sergeant York test program to establish his premise that this was one of the missions of test and evaluation. He likewise urged examinations of the fit of the system to the employment concept.

Dr. Dickinson emphasized the importance of training from two perspectives. He raised the issue of the role that training plays in preparation for test and the penalty the tester places on the system undergoing T&E, particularly operational test, when the operator is not adequately prepared to employ the system. He also contended that evaluating training was a part of the evaluation of a system and should not be overlooked in the final analysis of the results of T&E. Details will make or break. An important new capability could be rendered ineffective without adequate training, well-thought out tactics and techniques, and proper integration into the force. A comprehensive test and evaluation program will include an integrated consideration of all these aspects.

In addressing the strong push to use simulation to prepare for and to support operational test, **Dr. Dickinson** emphasized adequate resourcing of the necessary tools and the need to use effectively the entire gamut of available techniques—models to field test.

Dr. Dickinson's concluding remark on evaluation effectively summarized the challenge set forth by all the panel members. When considering the test and evaluation process:

LET'S MAKE SURE THE PROCESS IS COHERENT:

TEST AND EVALUATION WILL ONLY BE AS GOOD AS IT'S WEAKEST LINK!

WORKING GROUP I

EVALUATION AS A TOOL FOR TEST PLANNING AND IMPROVED TEST EXECUTION

Co-chairs: Dr. Darrell Collier
Director
US Army Training and Doctrine Command
Analysis Center, White Sands

Mr. Ray Jones
Manager, Operations Analysis
Advanced Development Engineering
Missile Systems Division
Rockwell International Corporation

Rapporteur: Ms. Barbara Toohill
MITRE Corporation

INTRODUCTION

Working Group I was charged with considering the proper role of evaluation in support of test planning and execution. The charter for the group is summarized below.

> Address the use of evaluation as a tool for test planning and improved test execution. Given the complexity of systems being tested and the environments in which they need to be tested, consider how we can use pre-test analysis to better plan and execute testing for the most efficient use of resources and the most insightful test results.

Note the key phrases concern supporting test planning and execution. The emphasis is on pre-test analysis and includes the need to both conserve resources and increase the effectiveness of the test results. This latter portion of the charter was taken to mean the insightfulness of both the test results and the resulting evaluation.

The approach taken in Working Group 1 activities aimed at synthesizing group member consensus on the assigned topic, and is portrayed in Figure 1.

FIGURE 1. WORKING GROUP I APPROACH

In initial group sessions on the first day, a general perspective was provided by

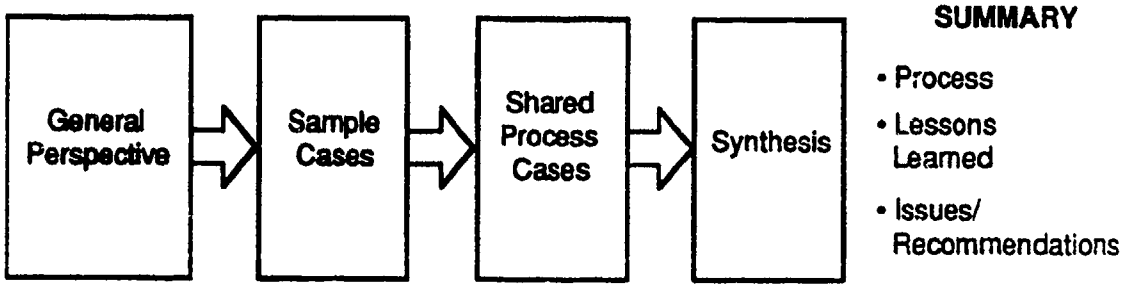


Figure 1. Working Group I Approach

group co-chairs. Evaluation for Test Planning and Test Execution was reviewed from a developmental and an operational test orientation. In addition, a review of process "flow" was provided to prepare group members for an assignment to develop processes, lessons learned, and issues based on their past experiences. Beginning the next morning, five specific cases were presented as below:

Topic and Presenter

"Improved Data Link (IDL) Tactical Weapon Simulator (TWS) Evaluation/Test Planning," Mr Ray Jones

"JTIDS Test Planning/Pre-Test Analysis," Dr Donald Van Arman

"Air Defense System Model-Test-Model," Dr Darrell Collier

"Model-Test-Model M1A2 EUTE," CPT Eugene Paulo

"Mission Level Measures, Requirements, and Concepts of Operations,"LTC Gerald Simnacher

After those more formal presentations, and based on the process "flow" approach presented earlier, several members of the group presented informal discussions of past related experiences as shown:

Topic and Presenter

"Evolving Weapon Separation Test and Evaluation at AEDC," Dr Ed Kraft

"Evaluation for Wind Tunnel Test Planning and Execution," Mr Russell Sorrells

"NWC Systems Testing and Evaluation Experiences," Dr Wildon Blackburn

"USMC OT&E Perspective," Maj Rick Reece

"Upfront Evaluation Structure for the TEMP," Mr. Hap Miller

After the background from general perspective presentations, sample cases, and shared process cases, the working group proceeded to the essential stage of consensus synthesis. The group was broken into four

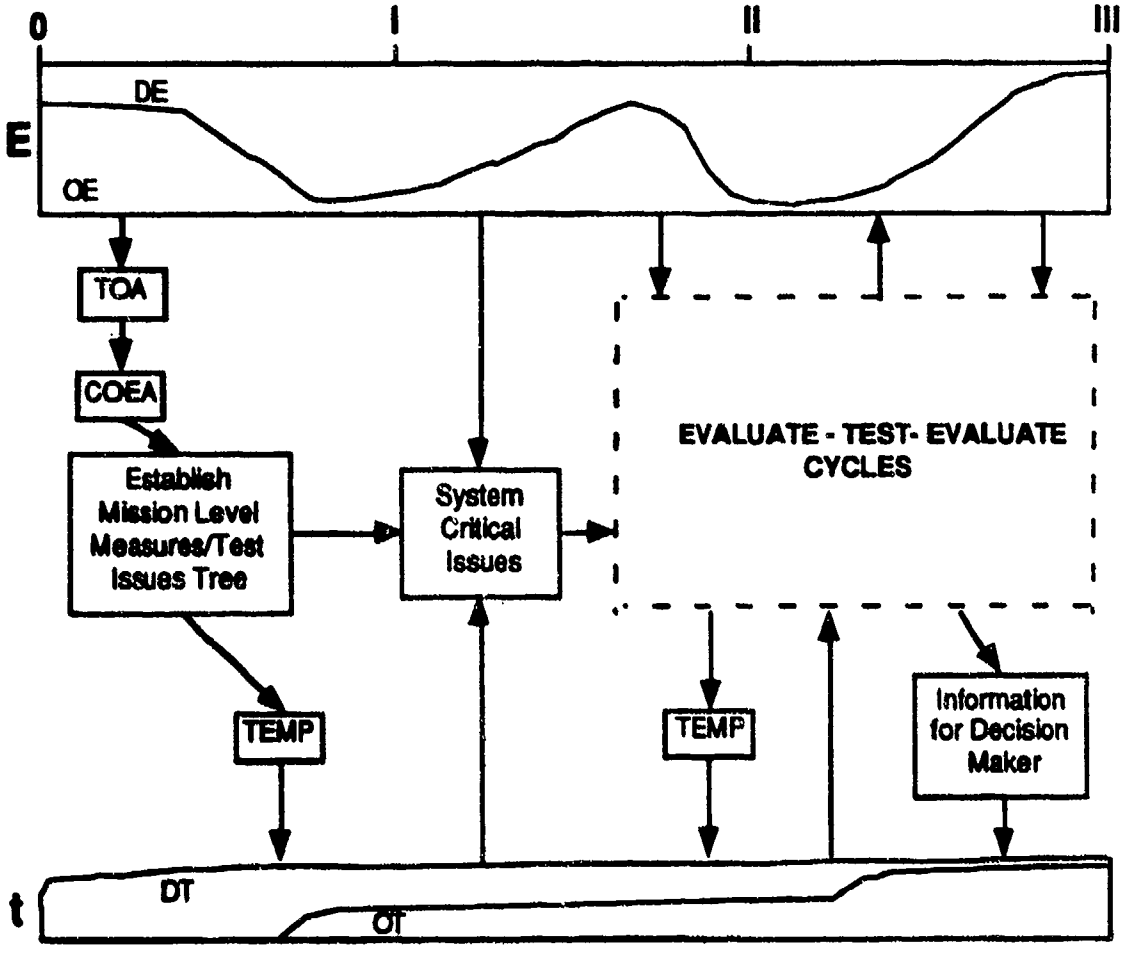


Figure 2. Evaluation for Test Planning/Execution Process

subgroups to synthesize group feelings in the following areas: Process, Benefits, Tools and Techniques, Issues and Recommendations. Outputs from the subgroups were reviewed by the entire working group.

EVALUATION FOR TEST PLANNING/EXECUTION PROCESS

After reviewing Working Group I discussions (and based on the broad based experiences of the group), a desirable process to better use evaluation for test planning and execution was developed. That process is portrayed at two levels: across system acquisition phases and at an iterative cyclical level repeated within those system acquisition phases.

In Figure 2 the broader Evaluation for Test Planning/Execution process is shown across the systems acquisition phases from Milestones 0 to III. At the top of the figure, evaluation activities are represented with a capital "E" to signify the increasing importance of effective evaluation. The total evaluation effort is comprised of two types: Developmental Evaluation (DE) emphasizing systems analysis type evaluation and Operational Evaluation (OE) addressing operational effectiveness type evaluation. Note that the relative mix of DE and OE varies through the systems acquisition phases. During early Concept Exploration and Definition (CE&D), effective OE is especially essential as the operational requirements are reviewed to establish viable system concepts. In later CE&D, DE becomes increasingly important as system concept designs mature. Generally, another emphasis on OE occurs prior to Milestone II, as operational effectiveness updates are needed to ensure system performance prior to the decision to proceed to Engineering and Manufacturing Development (E&MD). During E&MD, early emphasis returns to DE then back to OE as E&MD progresses to full-fledged operational testing. The testing portion of T&E activities is represented in the lower portion of Figure 2 by the lower case . Note that the level of testing activity, represented by a lower height initially indicates the reduced level of testing during early CE&D which is generally basic technology demonstration. OT activities begin at a lower level prior to Milestone I and continue at that reduced level until the later portion of E&MD when OT becomes dominant.

In the center area of Figure 2, several items are emphasized as especially critical to a successful evaluation for test process. Working Group I discussions highlight the need for an integrated evaluation plan including a mission level measure/test issue tree laying a framework relating system/subsystem/component technical characteristics to intermediate levels of measures going finally to top level system performance mission level measures. This type structure should be established early in the CE&D phase while preliminary trade-off analyses (TOA) and cost and operational effectiveness analyses (COEA) are being planned, coordinated, and performed. An example of a mission level measure tree is shown in Figure 3. When such an integrated performance evaluation approach is established early, it can very positively influence the Test and Evaluation Master Plan (TEMP) as well as aid in identifying system critical issues. With TEMP and system critical issues identified, a series of Evaluate-Test-Evaluate cycles begins, continuing through remaining system acquisition phases. During those iterative cycles, the TEMP would be updated at least once prior to Milestone II. And, based on those cycles at appropriate times in the acquisition phases, the desired output of all T&E activities—useful information (not just data)—would be provided to system acquisition decision makers.

EVALUATE-TEST-EVALUATE CYCLE

Figure 4 illustrates a generic Evaluate-Test-Evaluate (E-T-E) cycle through an iteration of Pretest Planning & Evaluation, Test Planning, and Test Evaluation. An E-T-E cycle begins with a Pre-Test Planning and Evaluation activity indicated by a triangle at the top of the figure.

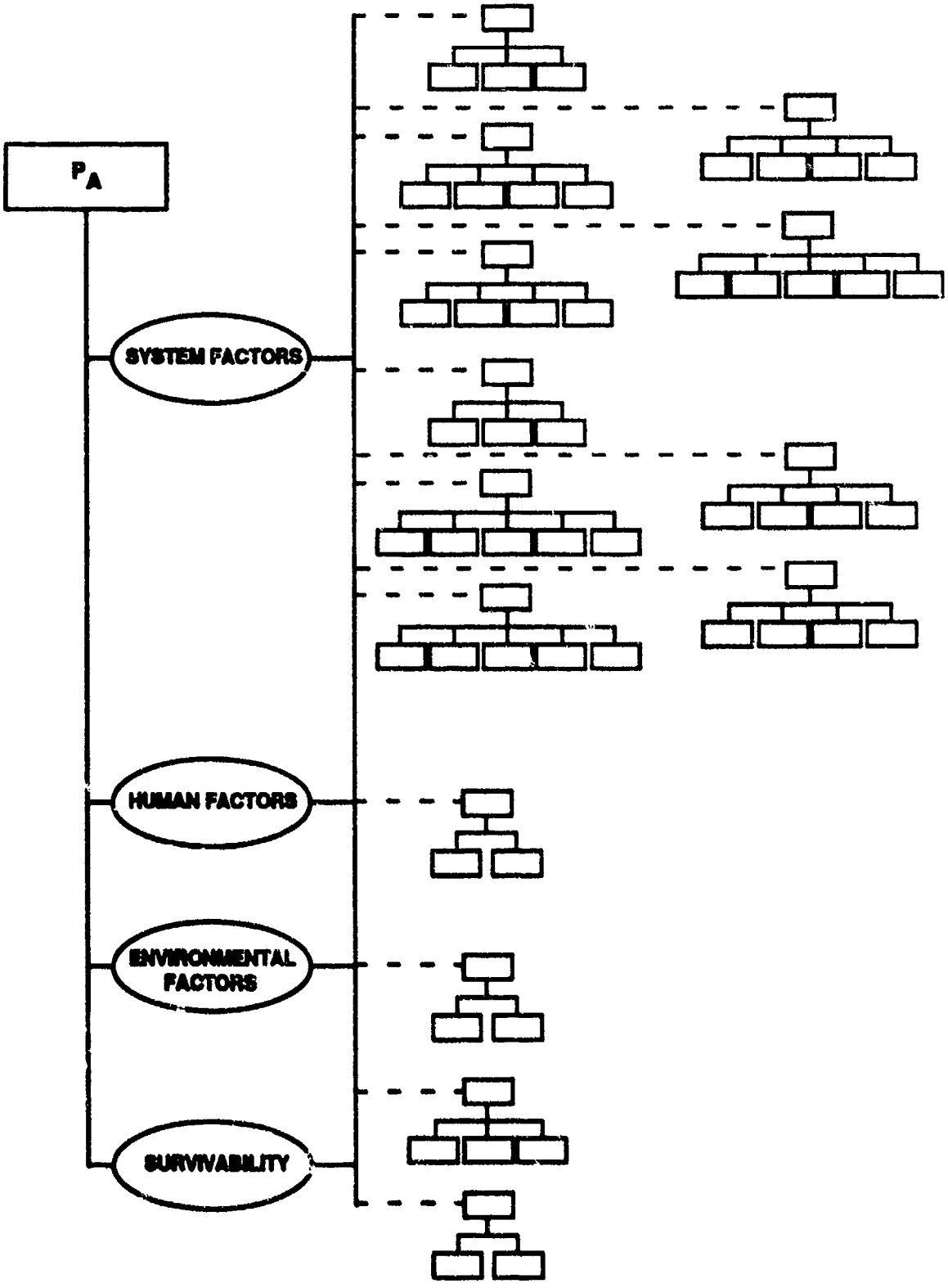


Figure 3. Example of Mission Level Measure Tree

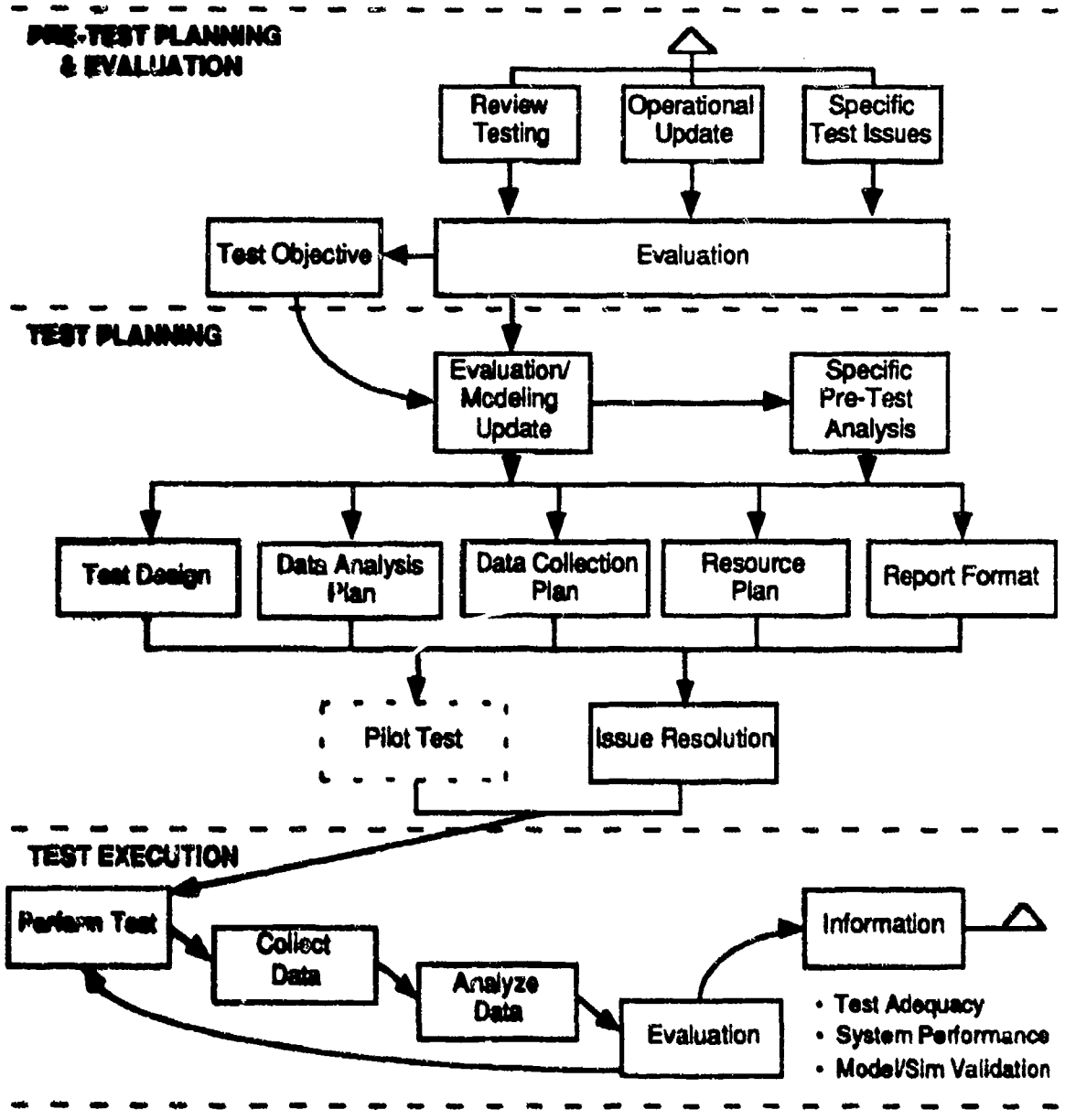


Figure 4. Evaluate - Test - Evaluate Cycle

Initially, post test results are reviewed, changes in operational requirements (target, threat, etc.) are noted, and specific test issues are highlighted. A preliminary evaluation relating to those background factors is necessary to clearly identify the test objective before moving to the Test Planning activity of the E-T-E cycle.

First in the Test Planning activity, models and evaluation techniques are scrutinized for applicability/required modification/update prior to performing specific pretest analysis. After that evaluation/modeling update and pre-flight analysis, specific test planning should proceed. It is essential to address all aspects of the planned test including Test (or Experimental) Design, Data Analysis, Data Collection, Resource (Facilities), and Report (Documentation) Format. Members of the working group felt strongly that test planning should address at least those aspects and that, in order to extract the maximum useful information, the test (or experiment) must be designed to support effective data collection and analysis. Rehearsal of data collection procedures has proven very useful. The availability of testing resources is best aided by early identification of particularly unique resources (targets, test hardware, telemetry, etc). Finally, if the test report format is established prior to the test, a much more efficient conversion to final report format is possible. One option identified as appropriate in many cases was a pilot test. In the pilot test, test assets and procedures are exercised to ensure that the experimental design supports the data collection/analysis and that results are readily transformable to the report format. The final activity under Test Planning is to resolve all identified issues prior to Test Execution.

That Test Execution effort is portrayed in the lower portion of the figure. Initial testing is performed yielding data which is collected and analyzed leading to an evaluation of test results. During Test Execution, it is essential for efficient testing that information from evaluation be fed directly back to Test Execution activities. In complex tests such as force-on-force operational tests, some of the evaluation tools (e.g., combat models) can assist in the analysis and interpretation of results. Generally, Test Execution includes iterations of test, collect data, analyze data, and evaluation. This allows follow-on test stages to gain efficiency from testing lessons learned. At appropriate times and after all data is analyzed and results evaluated, that essential output of useful information is provided to program decision makers and should address testing adequacy, system performance, and model/simulation validation.

EVALUATION FOR TEST PLANNING/EXECUTION PROCESS-GENERAL OBSERVATIONS

Working Group I findings relating to the Evaluation for Test Planning/Execution Process are summarized below:

- A system Mission Level Measures (MLM) Tree is strongly recommended early in the CE&D phase in conjunction with drafting the initial TEMP.

- Based on review of the MLM Tree, unique test facility or resource requirements should be identified.

- The overall Evaluation for Test Planning/Execution process contains iterative looping of Evaluate-Test-Evaluate cycles.

- The generic Evaluate-Test-Evaluate cycle includes three primary phases of activity: Pretest Planning and Evaluation, Test Planning, and Test Execution.

- Evaluation yielding clear test objectives is required prior to even beginning Test Planning.

- Test Planning must address Test (or Experimental) Design, Data Analysis, Data Collection, Resources (or Facilities), and output Report (Documentation) Format.
- Testing prior to actual Test Execution is desirable in most cases.

- Staged Test Execution allows feedback of evaluation from previous stages and provides increased efficiency.

- The desired output of the Evaluation for Test Planning/Execution process is useful information addressing test adequacy, system performance and model/simulation validation.

MAJOR BENEFITS

Several benefits are gained from conducting evaluation prior to, during, and after the test. Reasons for utilizing the Evaluate-Test-Evaluate cycle are highlighted below.

- To elaborate system requirements

- To conduct "smarter" tests

  - Achieve better definition during test design phase

  - Achieve better economy and efficiencies during test planning and conduct phases

- To enhance quality and credibility of T&E results

A pre-test evaluation allows a significant opportunity to further elaborate system requirements. Such an effort provides a sanity-check regarding realistic expectations for system performance as well as a better understanding and characterization of projected system employment techniques. Analysis at this point helps illuminate trade-offs of lower-level parameters while still meeting macro-level mission requirements. The results of such efforts determine gaps in our knowledge and establish baseline data and analyses to guide future evaluation. Overall, pre-test evaluation provides information to the decision maker to assist in system risk management.

Pre-test evaluation fosters the conduct of "smarter" tests. It allows the evaluator to achieve better definition during the test design phase and encourages integration and accumulation of knowledge across different phases. It focuses tests on filling gaps in our knowledge and reduces the chance of collecting unnecessary data. It is easier to clarify test objectives and provides a rational method to reconcile different perceptions of issues for the next test phase. Such an effort is well suited to better relate MOEs and critical operational and developmental issues.

A strong emphasis on pre-test evaluation assists in identifying threats to the validity of the test (i.e., identifies those things which could give the same result but for a wrong reason) as well as providing a paradigm for the final analysis, enabling better explanations of the "whys." It helps define relevant test conditions and scenarios

Build Conceptual models to determine:
- -what questions you are trying to answer
- -the measures that should be used
- -the criteria that will be used to make decisions

*Review of COEAs
*Modeling and simulation
*Interdisciplinary input & review
*Decision sciences
- -multi-attribute utility theory
- -analytical hierarchy process
- -delphi techniques

*Expert systems / AI / decision trees

Chart 1. Tools and Techniques (I)

(threat and environment) and resource needs (range, personnel, etc.). Pre-test evaluation assists in defining qualitative and quantitative measures and appropriate statistical methods by providing hypothetical test results and aids in the proper choice of test parameters such as sample size.

Perhaps the strongest reason for conducting pre-test evaluation is to project anticipated T&E information (types and levels of evidence) to reduce surprises in what might be known or not known at the end of the T&E. Indeed, this allows complete draft evaluation to be performed emphasizing everything from analytical techniques to planning how the T&E information will be presented (i.e., how the information about system capabilities will be conveyed in an accurate and convincing manner). The use of this process contributes to better economy and efficiencies during test planning and execution phases and allows us to reduce the scope of OT by identifying those test events that produce no additional value added information. It supports determining sample size based on desired levels of confidence (statistical vs. meaningful significance) and helps identify evaluation tools needed during and after test.

Pre-test analysis—particularly with a heavy reliance of simulation of tests—provides significant and meaningful information for T&E risk management by producing a better understanding of the trade-offs between practical test plans and the level of evidence desired for findings. It allows us to better handle test and analysis costs.

The Evaluate-Test-Evaluate methodology allows us to enhance the quality and credibility of test and evaluation

results. Through this process we identify (and determine the importance of) assumptions and possible biases. We ensure appropriate data are collected to avoid the "oops, we forgot that" embarrassment. This assures the reasonable interpretation of test results and -foremost - fosters communication and coordination up the "food chain" leading to better understanding of what information is needed for decision making.

TOOLS AND TECHNIQUES

No matter how attractive Evaluate-Test-Evaluate is as a process, it is of little use if the tools and techniques do not exist with which to implement it. In evaluation prior to testing, one important step is to build a conceptual model of both the system and the potential test. As shown in Chart 1, this modeling process helps us determine what questions we are answering in evaluating and ultimately testing the system. As we understand the system, evaluation measures are determined and criteria established to enable us to make decisions.

There are several tools and techniques which support us in this process. However, there is no "cookbook" for the effort, but rather a need to consider a wide array of tools and techniques. First, we review all that has gone before, with particular emphasis on the Cost and Operational Effectiveness Analysis to set the stage for providing initial direction and scope to conceptual modeling. Modeling and simulation tools and techniques provide powerful vehicles for investigating the system issues and determining what we do and do not know about the requirements of the system and its proposed operational use. Of particular interest are models and simulations used in prior analyses such as the COEA and any cost-performance trade-off analyses. It is critical during this process that a team composed of representatives from several disciplines be involved in the modeling process. Experts in the system and technology domains along with threat analysts help define aspects of the system which have risk potential. Operational experts assist in identifying areas where operational or environmental factors can introduce significant uncertainty. The key is to obtain the widest possible advice and to benefit from the broadest experience.

Early in the evaluation process there may be many soft areas both in understanding of the system and of the subsequent decision process. In the conceptual process, consideration is given to techniques suitable for handling more subjective aspects of the system and its performance. Many techniques help in the conceptualization of the decision process and in the preliminary evaluation of risks and uncertainties. Modeling of subsequent evaluation is also considered through techniques such as expert systems or decision trees.

The key is to not emphasize a particular set of techniques, but rather to examine a set to select the most appropriate for the purpose at hand. The utility of the evaluate-test-evaluate process is also impacted by the availability of tools and techniques to support quantitatively the steps shown. These areas are key to improving the quality of evaluation through better designed, more focused testing.

Use analytical methods to :
- -determine which MOEs drive the evaluation
- -identify and resolve problems (e.g. data collection and instrumentation)
- -practice testing techniques
- -draft independent evaluation report

- *Modeling and Simulation
- *Sensitivity / trade-off analyses
- *Techniques appropriate to the test
  - -methods for small sample sizes
  - -non-parametric statistics
  - -experimental design

Chart 2. Tools and Techniques (II)

The development and use of system/test environment models and simulations are an essential ingredient of the methodology. Such tools not only embody the latest information concerning the system performance and interaction with the environment but allow exploration of the interaction of the system with that environment. All types of models and simulations are used. In the operational arena these range from interactive or closed analytical combat simulations (such as COEAs) to man-in-the-loop flight or crew simulators (such as SIMNET). In the developmental arena, these models and simulations include detailed physics and engineering models of the system and its interaction with its environment such as aerodynamic models of ordnance release or hardware-in-the-loop hybrid simulations of prototype systems.

To be most useful, tools must handle early system concept information and support exploration of the performance envelope (identification of the "driving factors") through sensitivity analyses. Tools which will also support trade-off analyses are of most use during this phase since it is this exploration which generates the understanding. The evidence is that many such models and simulations exist with direct applicability to pre-test evaluation.

Modeling of the test also contributes to an understanding of potential test outcomes and opens up the selection to a wide variety of tools which contribute to efficient testing and may be more appropriate to the type of data expected. In particular, one can explore the appropriate experimental design or best technique to evaluate the data.

Consider a single, integrated evaluation plan vice the current TEMP

Pros:

- -emphasizes evaluation
- -allows choice of most appropriate technique
- -clear road map for test requirements

Cons:

- -who's in charge?
- -proper break-out of OT / DT?

Chart 3. Matters for Consideration

The result of appropriate use of these techniques is the ability to practice the test and draft the evaluation before committing to the expenditure of test resources. This is key to ensuring that surprises from the test are due to unexpected performance of the system under test - and not due to a lack of thought about the test conditions.

ISSUES

Discussions by the working group raised several issues concerning evaluation for test planning. These are listed below.

- Is the current TEMP structure an adequate roadmap for evaluation?
- Are DT/OT properly coordinated?
- Is the COEA properly linked to the evaluation?
- Is the "BEST" technique to support each evaluation issue identified?
- Are resources to support the evaluation addressed early enough?

There is significant concern that the usual format and structure of the Test and Evaluation Master Plan is not conducive to providing a proper roadmap for the evaluation process. In fact, there is a wide spread perception that the TEMP is too test oriented and it should be changed to concentrate on evaluation or prefaced with an evaluation document. During group discussions, reference was made to separate chapters for DT&E and OT&E and the general perception created by the difference between the depth and detail of the respective efforts. Some discussion occurred relative to the "ownership" of the TEMP by the Program Manager and the subsequent impact on the form and content of the document—but the general consensus was that while that might contribute to a lack of evaluation perspective, it did not cause it. The focus on evaluation could be improved

by including specific material referencing the COEA and T&E linkage process.

The working group generally agreed that DT and OT should be coordinated so that information from developmental testing was available as a coherent part of the evaluation. This was particularly important to avoid unnecessary duplication and to ensure that subsequent OT was planned on the basis of the latest available information. Most members of the working group, without regard to service orientation or experience, agreed that neither DT nor OT maintained the proper cognizance of the other efforts. It was generally agreed that this was a result of the apparently independent development of the DT/OT plans. Separate chapters in the TEMP were symptomatic of this divergence.

There appears to be little evidence of the Cost and Operational Effectiveness Analysis playing a significant role in the planning for evaluation. The COEA, it is generally agreed, is an excellent starting point for development of an evaluation plan, but there is little experience that this is occurring. Most participants were aware of benefits of the linkage of the COEA to evaluation, but were unaware of specific instances of application.

A perception of test and evaluation planning most commonly portrayed is of a tendency to reach for the "usual" or traditional techniques instead of an objective search for the "best" tools to support the evaluation process. A symptom of this is the perception that the evaluation plan is often "business as usual" instead of a focused effort driven by the issues at hand.

This lack of perceived coherent planning is believed to contribute to a lack of early identification of evaluation resources.

RECOMMENDATION

Based on discussions and considerations of the working group as presented here, Working Group 1 recommends that consideration be given to the matter shown on Chart 3. We believe that at the current time, there is mixed emphasis in the initial test planning stages. Even the title of the driving document is symptomatic of the overall issue—that is, the TEMP—or the TEST and Evaluation Master Plan. This orientation encourages emphasis on planning for test and specifically encourages in initial planning stages a separation of DT from OT. To achieve and maintain the correct emphasis on the primacy of the evaluation, it would help to change the name and structure of the document from which all subsequent planning must flow.

Serious consideration should be given to replacing the TEMP with a document which provides a single, integrated EVALUATION Plan. This document should be considered as the capstone, a living document in which all evaluation requirements are coordinated and from which all activity in support of the evaluation is derived throughout the life of the program.

Such a change would accomplish several things. First, it would clearly emphasize the plan for evaluation as the first and foremost requirement in the test planning/execution process. Second, it would shift the emphasis from always

assuming that the supporting activity is test related to consideration of the most appropriate technique—which may be simulation based. And third, it would provide a clear road map for test requirements—based on the needs of the evaluation.

Such a change is not without problems. There is always an issue of "who's in charge?" of such a document. The obvious choice is the operational evaluator, but involvement of the developmental side is important. Similarly, there is the possibility that such an arrangement might lead to insufficient emphasis on developmental testing. Overall balance and completeness is the driving issue. The question is how to achieve it. Working Group 1 perceives that increased emphasis on an integrated evaluation plan is a driver needed to improve test planning and execution.

WORKING GROUP I PARTICIPANTS:

NAME	ORGANIZATION
Professor Robert E. Ball	Naval Postgraduate School
Mr. Wildon C. Blackburn	Naval Weapons Center
Dr. Darell W. Collier	US Army TRAC-WSMR
Mr. William C. Edwards	National Security Agency
Dr. Harvey J. Finberg	US GAO
Ms. Christine A. Fossett	US GAO
Mr. Richard E. Garvey, Jr.	BBN Systems & Technologies
Mr. Raymond A. Jones	Rockwell International
Dr. Edward M. Craft	Calspan Corp/AEDC Division
Mr. Philip R. Landweer	BDM International
Mr. Richard S. Miller	Institute for Defense Analyses
Mr. Charles W. Mullis	STEWS-NR-PR
Mr. William A. Niemeyer	US AMSAA
Mr. Richard W. Pace	Naval Air Test Center
CPT Eugene P. Paulo	US Army TRAC-MTRY
LTC Michael C. Proctor	US Army TRAC-MTRY
MAJ Rick L. Reece	H&S BHn, MCOTEA
Mr. Charles A. Reveal	6545th Test Group/ENAT
Lt. Col. Gerald Simnacher	HQ AFOTEC/ST
Mr. Russell B. Sorrells	USAF/AEDC/DOF
Ms. Barbara G. Toohill	The MITRE Corporation
COL Michael T. Toole	SDIO/TNE
Mr. Erling D. Tranby	Naval Ship Weapon Systems Engineering Sta
Dr. Donald Van Arman	The MITRE Corporation
Mr. Allan E. White	Kapos Associates
LTC Stephen D. Williams	US Army War College
LT Cheryl K. Worlein	OPTEVFOR

WORKING GROUP II

EVALUATION FRAMEWORK TO CLOSE THE LOOP IN THE ACQUISITION PROCESS

Co-chairs: Dr. Ernest Seglie
Science Advisor
Office of the Director of Operational Test and Evaluation
Office of the Secretary of Defense

Mr. John Marshall
Deputy Program Manager, Unmanned Vehicles
Naval Air Test Center

Rapporteur: Mr. Theodore Bean
MITRE Corporation

The charter of this working group was to "address the use of an evaluation framework to close the loop in the acquisition process." It was to "examine a possible architecture to link mission area analysis for requirements definition, cost and operational effectiveness analysis, and analysis of the results of test and evaluation."

The working group included over 30 individuals in the military (captain to colonel), civilian government (GS-12 to SES), and private industry. Most are evaluators, have had that responsibility in the past, or are currently consultants to evaluators. Each had the opportunity to identify areas that could be improved and suggest improvements. Thirty-six specific problem areas were identified. At the time of these discussions, early October 1991, the new series of acquisition directives, instructions, and manuals (the "5000 Series") has not had a chance to demonstrate whether they will significantly change the practice and interaction of the user, developer, and evaluator. On the other hand, there is evidence to suggest that programs are attempting to skip steps that have a significant role in ensuring good evaluations. The working group believes there are ways to improve the evaluations further in order to better serve the decision maker.

Fundamental to the group's view is the recognition that analysis should underlie the Operational Requirements Document (ORD), analysis should underlie program evaluation, and analysis should underlie the operational evaluation. If these analyses form a consistent set, each evaluation will be stronger than it is today. The immediate need is to strengthen the linkage between elements of DoD evaluations; communications between evaluators and users; and the requirements development process. Examples of the problem abound: the requirements process was described

during the plenary session as sometimes "flakey;" the Joint Requirements Oversight Council frequently is asked to change requirements just before full rate production decisions. It was felt improvements are possible. For example, more focused programs and evaluations would result if all evaluators were involved in the development of the Cost and Operational Effectiveness Analysis (COEA) and the TEMP, and that coordination among evaluators and users could increase the chance of getting a statement of precise testable requirements which is considered a must.

The acquisition process is complicated, extended over time, and has many players. To be effective, the acquisition process must have integrity. The evaluations that occur must form a whole, with all the pieces consistent. The working group felt that the integrity of the process can be significantly improved, and improvement will bring a more productive process of evaluation.

The integrity, oneness, of the evaluation process can be improved by better discipline and communication: discipline within the hierarchy of evaluations and communication vertically and horizontally among contributors to the process.

THE HIERARCHY OF EVALUATIONS

The new DoD 5000 Series Acquisition Directives contain a hierarchy of evaluations that are specified as products to be delivered at milestones. What is missing is a procedure to ensure a logical relation between these products, and second, a standard for that connectivity. The products that must be better linked are the Mission Area Analysis (MAA), the Mission Need Statement (MNS), the Operational Requirements Document (ORD), the Cost and Operational Effectiveness Analysis (COEA), and the Test and Evaluation Master Plan (TEMP). ORDs, TEMPs, and COEAs appear first at Milestone I and are updated at Milestone II. How the documents are related to each other is not specified, and no one is responsible to ensure traceability from one document to the next. Consider the MAA and the MNS. Both should assess a military capability, but the rationale for conclusions is not required. At present, only casual links exist between them. The logic and rationale which lead from one to the next tend to be weak or unsubstantiated. In the process of analyzing the mission area and the military need there is a need to identify what measures of effectiveness were or should be used. Later the COEA must identify measures of effectiveness that may or may not be related to those used, but hidden, in the MAA. Later still, the DT and OT face the same problem. The primary cause of the lack of traceability is the diversity of organizations (with differing levels of experience and training) responsible for the different evaluation products. A logical relationship between the MNS, ORD, COEA, and evaluation MOEs is hard to achieve unless the organizations responsible for developing each have a better working relationship.

The lack of linkage between evaluations is best illustrated by how seldom the two evaluation documents available to OSD refer to each other. The TEMP should include or reference the COEA measures of effectiveness and suitability at each TEMP update. Both should identify and provide rationale for changes in scenario/threats and in the related requirements. All should be

logically traceable from the Mission Area Analysis.

A specific example of the disconnects that can exist is in the area of scenario development. Military scenarios form the context of all evaluation processes, analytic, and testing evaluations. Scenarios are missing. Common Mission Scenario Models should be used in COEAs for all those types of systems that interface in the same mission scenarios. The spectrum of scenarios needed start at the highest level: J-8 exemplar scenarios which should reflect the Defence Planning Guidance. The traceability down to scenarios for the COEA and for the operational tests needs to be clear. As one participant stated, "I am convinced that the little scenarios I am given to work are delivered by the stork, because I can't find any parents for them."

The greatest area of need is for a more well thought out requirements process that includes cost-benefit and trade-off analysis from the COEA/MAA process and that involves the DT and OT agencies. The MNS needs to be strengthened by better analysis at the start of the process. The COEA should be complete in the sense of including sensitivity analysis and all relevant, affordable options. Not all criteria need to be quantifiable; however, some value should be evident to the intangibles. Further, the criteria that are quantifiable need not be expressed as a single value, rather, a range of values seems more appropriate. This range helps one to answer the question of the relative goodness of a system performing at varying levels of effectiveness which can be ascertained via sensitivity analysis. With these, decision makers should understand what they are paying for, how much various capabilities cost, and the testers should understand what is important.

A first step is to ensure that all the documentation, in approved form, is available at the milestone. If the TEMP or COEA has shortfalls, there will be no approval for the next phase. OSD standards need to be established and followed. Not enough is known early enough about the evaluation criteria. All evaluators should agree that if these criteria are met, then the deficiency is satisfied. Since the deficiency is written in non-system-specific terms, these MOEs can be developed early in the MNS and translated to system specific terms latter in the ORD, COEA, and TEMP.

There must be a mechanism for continuously updating the COEA based on changing threats and/or missions scenarios, test results identifying limitations in system capability and sensitivity of mission success to these limitations, and changes in cost. In addition, all programs should go through the COEA process. Currently only a few programs go to OSD. With limited resources, the Services are inclined to short-cut the process.

A single evaluation framework is needed. If necessary, a Master Evaluation Plan might be appropriate. An alternative mechanism for developing the needed linkage and communication is a Test and Evaluation Integration Working Group which includes all the evaluators, COEA, DT&E, OT&E, and the user.

Formation of an evaluation oversight group populated by the operational and technical evaluators, the requirement generators, and the leader of the COEA, to oversee the life cycle evaluation of weapons

systems prior to Milestone 0, has the potential to overcome this shortfall. The group would assure that methodologies, rationales, and evaluations are consistent throughout the life cycle. One Service, the Marine Corps, has already begun forming such COEA Study Advisory Groups.

TOOLS TO IMPROVE EVALUATION

Lack of communication is a major problem in the evaluation process. The basis for communication ought to be a standard architecture that addresses effectiveness and suitability evaluation. Better communications are preferred to more bureaucracy: first get the right people together, then give them the right tools.

The COEA Study Advisory Group already mentioned is one possible response. It was developed because no single organization in the Service had enough people in one place to do the study itself. The requirements people, developers, an intelligence representative, and a headquarters representative need to agree on MOEs with the testers. The committee approach also allows independence. The purpose of this "lead-off summit" is to get improved communications and understanding of:

- mission requirements and scenarios
- technology and engineering options vs cost
- measures of merit vs test limitations.

Rigorous work to support Milestone 0 and I is needed. Since Milestone 0 usually does not involve a large commitment of funds, there is a tendency to let things slide. The early stages are exactly when everything can be set on the right track or left to wander. While most of the responsibility must fall to the Services at this early stage, there should be coordination between the OSD evaluators, PA&E, DOT&E, and DT&E with JROC representative. On the whole there should be better demonstration by OSD of a willingness to work with the Services early in the process to help get the job done.

The above suggests who should be involved, but next must come the tools they use. One tool that has promise is a model-test-model procedure, or as Working Group I suggested, an evaluate-test-evaluate cycle. Whatever the name, the procedure can form an evaluation framework for system acquisition, tactics, and concepts development. It can be cheaper than additional testing, and can improve the cost effectiveness of current testing by "testing the test" before it goes to the field. The model-test-model approach should begin with a well understood and agreed-upon evaluation criteria.

The COEA can be the tool for communications in this model-test-model framework. The models used in the COEA are the first to represent, at one time, the proposed system, the concept of operations, the required interoperability, and the relevant scenarios. If the model used in the COEA is run with the OT scenario, it could also be the first representation of the test. The models used in the COEA have a further potential advantage because engineering trade-offs are made on the basis of cost-effectiveness trades. One practitioner outside DoD noted that his agency uses the

COEA to overcome the fact that the user and developer have different languages. He locks the user and the engineers together to write the COEA. In that way the engineer gets to really understand what the user wants to accomplish, and the user to understands better which of his desires is costing the most. The COEA can also increase the credibility of the evaluations by "testing the test."

Analysis underlies the ORD, program evaluation, and operational evaluation. The COEA can and should serve all three purposes if it is constructed to do so and is perceived as a critical document. At a minimum, the COEA and TEMP should use parallel structures: scenarios, threats, measures of effectiveness/ suitability.

PEOPLE AND RESOURCES

Additional training is needed to assure that the analytic and acquisition community understand the standards for "evaluation framework."

- A working group should be formed that can foster a COEA capability.
- Open the doors to informal discussion. COEA evaluators who must update the COEA based on test results should be encouraged to see the tests.
- The COEA analysts should participate in test integration working groups.
- All of this must be done with the recognition that the resources for people to do the analysis are becoming more scarce.

SUMMARY

In summary, the lack of connective tissue, links, logic, and rationale in the hierarchy of evaluation activities puts the decision maker in an awkward position. As one participant said, "I feel my boss's decision to let programs proceed is really a measure of his jumping ability. After asking reasonable questions to which the documentation provides no answer and for which the logic is missing, he has to make a leap of faith." The suggestions here are to insist that the logic be explicit and consistent; get the various evaluators together with the user to explicitly ensure consistency, and use some form of an iterative model-evaluate-test-evaluate cycle to foster a coherent analysis and evaluation.

WORKING GROUP II PARTICIPANTS:

NAME	ORGANIZATION
Mr. Theodore T. Bean	The MITRE Corporation
Mr. Tony R. Bellach	USAF/Office of Aerospace Studies
Ms. Linda J. Bors	HQ SAC/XPSM
Mr. Wilbert J. Brooks, Jr.	US AMSAA
Mr. Thomas P. Christie	Institute for Defense Analyses
Lt. Col. Allan L. Collie	HQ AFOTEC/XPX
Dr. Gary C. Comfort	Institute for Defense Analyses
Col. Donald M. Douglas	HQ AFOTEC/OA
Mr. Daniel J. Feighery	Institute for Defense Analyses
Mr. Ronald A. Gustafson	HQ AFOTEC/CNP
Dr. Robert G. Hinkle	Office, DUSA(OR)
Capt. K. D. Jenkins	H&S Bn, MCOTEA
Mr. W. Donald Johnson	US AMSAA
Mr. Sidney Kissin	National Security Agency
Ms. Theresa R. Kocher	HQ TECOM
Mr. Paul M. Lang	SRS Technologies
Ms. Ann M Lubas	MCRDAC
Mr. John A. Marshall	Naval Air Test Center
Mr. Gregory K. McGill	AFOTEC/ST
LTC L. James Merryman	Joint Staff, J-8
Dr. James J. Metzger	ODCSOPS, HQDA
Mr. Robert J. Meyer	Naval Weapons Center
Mr. Donald E. Mixon	The MITRE Corporation
Ms. Beverly Nichols	US Army Air Defense Artillery Center
Mr. Dale L. Paquette	Naval Underwater Systems Center
Mr. John M. Reeves	SAIC
Mr. Raymond W. Reig	Defense Systems Management College
Mr. Steve A. Schroffel	US Immigration and Naturalization
Dr. Ernest A. Seglie	OSD(DOT&E)
MAJ Cyrus J. Staniec	OASD(PA&E)
Mr. Edward G. Stauch	US Army TECOM
Lt. Col. Dewey D. Tucker	OASD(PA&E)
Ms. Sylvia L. Waller	Institute for Defense Analyses
Dr. Marion L. Williams	HQ AFOTEC

WORKING GROUP III

EVALUATION DURING DEVELOPMENT TESTING TO REDUCE RISKS

Co-chairs: Mr. Irv Boyles
Close Combat Evaluation Directorate
US Army Operational Test and Evaluation Command

Mr. C. W. (Chuck) Zang
BDM International, Inc.

Rapporteur: Mr. Sidney Polk
MITRE Corporation

Working Group III was tasked to address the use of evaluation during development testing and evaluation (DT&E) to reduce inherent risks during development, and to prevent unfortunate surprises in operational test and evaluation (OT&E). This working group was well suited to the task being composed of an outstanding representation of development testers and evaluators (government and contractors), operational testers and evaluators, and personnel from the Office of the Director, Test and Evaluation.

OVERVIEW OF APPROACH

Working Group III accomplished this task in the following order:

Step 1: Established who evaluators are, and their relationship with testers today.

Step 2: Summarized typical functions accomplished by development testers contrasted with evaluators.

Step 3: Identified types of risks inherent in DT&E.

Step 4: Identified opportunities and constraints expected within the five years for the T&E community from external sources; i.e., "external conditioners."

Step 5: Identified strengths and weaknesses within the

T&E community from which to respond to the external conditioners; i.e., the "internal conditioners."

Step 6: Given the external conditioners, and capitalizing on strengths while addressing the weaknesses of the T&E community, a "vision" was developed to express how we would like to see the development evaluator in five years.

Step 7: From the vision, strategic actions (recommendations) were identified to achieve the elements of the vision.

STEP 1

Question: Who are evaluators vs. testers today, and who drives testing?

Today, among military organizations and contractors, there is no consistency in either the distinction or organizational relationships between "tester" and "evaluator." The following general situations occur:

Situation 1: The tester and the evaluator are the same people. This usually occurs because of limited resources and "small" test programs.

Situation 2: Testers are different people from the evaluators, but are in the same organization. The supporting position for this situation was an allowance for specialized skills of testers and evaluators, but a single management responsible for products.

Situation 3: Testers and evaluators are different people reporting to different organizations. This situation allows for specialized skills of tester and evaluator, "independence" to assess validity of testing, and testing at numerous sites (ranges, facilities) with a single evaluation report.

In answer to "who, today, drives what testing is performed?" there was no consistency among defense organizations and contractors. Replies included the following:

- The tester drives the testing.
- The evaluator drives what testing is performed.
- The program manager drives what testing is performed (he normally pays for it).
- In DT&E, what will be tested will be that to be required during OT&E.
- The "user" or operating command to gain the new capability drives what testing is performed.
- OSD, particularly DOT&E or DDDRE(T&E), drives what is tested through approval or disapproval of the T&E Master Plan (TEMP), and/or the test concept/design.
- Contractor(s) may drive testing.

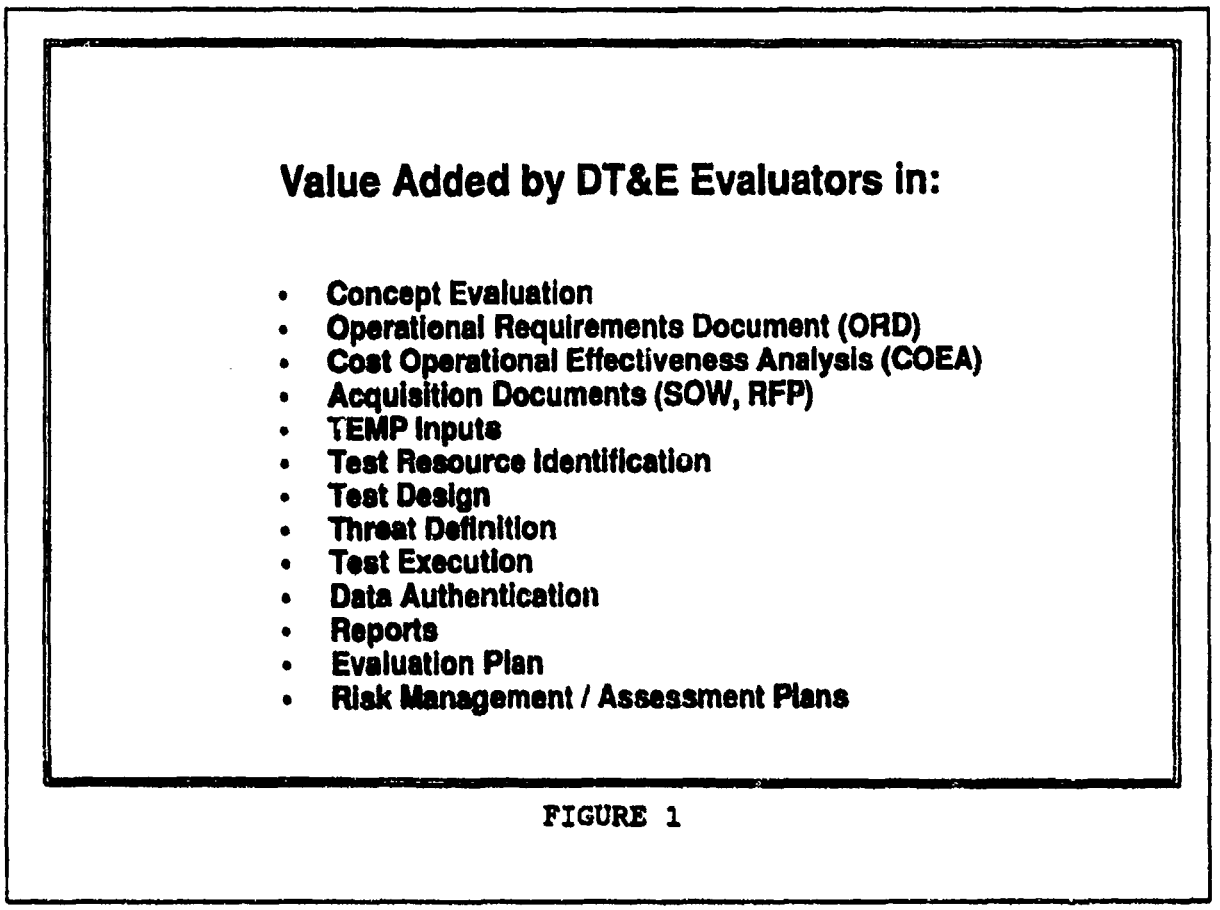


FIGURE 1

Indeed combinations of all these may be responsible for driving types and scope of development testing.

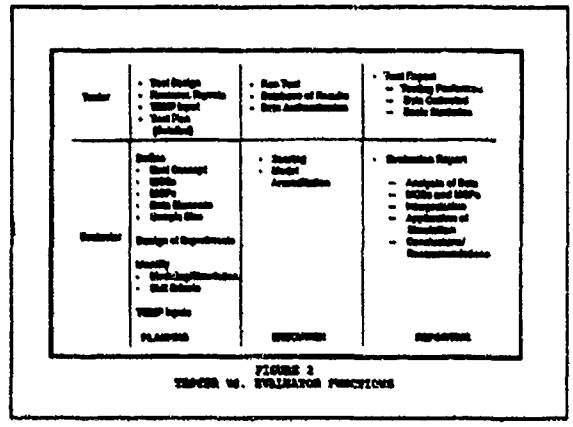
STEP 2

Question: What functions generally distinguish "evaluators" from "testers"?

This question was approached by first determining whether testers and/or evaluators provided or could provide any value added in the development acquisition process for the events listed in Figure 1. There was consensus that testers and evaluators do/should have a role in each of these events.

There was consensus in Working Group III that the functions depicted in Figure 2 are those most often associated with testers vs evaluators. These functions are broken out according to three phases of participation: test planning, test execution, and reporting results. These functions seemed to apply regardless of whether DT&E or OT&E was being considered. For some functions, both a tester and an evaluator participate from their respective roles.

Another way of looking at this distinction was to consider the "level" of data they generate. There was consensus that Figure 3 represented the full spectrum



of data produced during T&E, with testers generating levels 1 through 3 or 4, while evaluators generate level 4 through 7.

Figure 4 represented the most prevalent risks associated with DT&E. What the DT&E "evaluator" could do to reduce these risks became the focus of the Working Group III.

STEP 3

Question: What types of risks occur during DT&E?

To answer this question, Working Group III identified four primary purposes of DT&E:

1. Determine compliance with contract specifications and goals.

2. Isolate problems — assist in the development.

3. Certify that a system is ready for OT&E

4. Provide inputs for the cost and operational effectiveness analysis (COEA).

In this light, there was consensus that limitations in the items summarized in

DATA LEVELS	
LEVEL 1	RAW DATA - ORIGINAL FORM
LEVEL 2	ANNOTATED DATA - CONSOLIDATED RECORDS
LEVEL 3	AUTHENTICATED DATA - VERIFIED
LEVEL 4	DESCRIPTIVE DATA - SUMMARY STATISTICS
LEVEL 5	INFERENCES - PREPLANNED ANALYSIS
LEVEL 6	EXTENDED ANALYSIS - EXTRAPOLATION, MODELING
LEVEL 7	EVALUATIVE CONCLUSIONS - MILITARY JUDGEMENT

FIGURE 3

STEP 4

Question: What actions, events, and trends, over which the T&E community has no control, do we see coming that are likely to impact the T&E community? Do we consider them as opportunities or constraints?

After brainstorming, there was consensus within Working Group III that the items listed as "External Conditioners" in the following represent major influences expected in the next five years that will or can impact the T&E community. Some of these are identified as "opportunities" in that some aspect could "fall out" as a benefit, or with some initiative, could be made to benefit. Others are identified as "constraints" in that they are likely to force changes or limitations to the T&E community:

EXTERNAL CONDITIONERS (BEYOND CONTROL OF THE T&E COMMUNITY)

OPPORTUNITIES:

- Event Driven Schedules
- Increased Importance of Evaluation
- Models and Simulations
- Less Tolerance for Risks
- Exit Criteria
- Technology Demonstrations/Prototypes
- OSD Initiatives

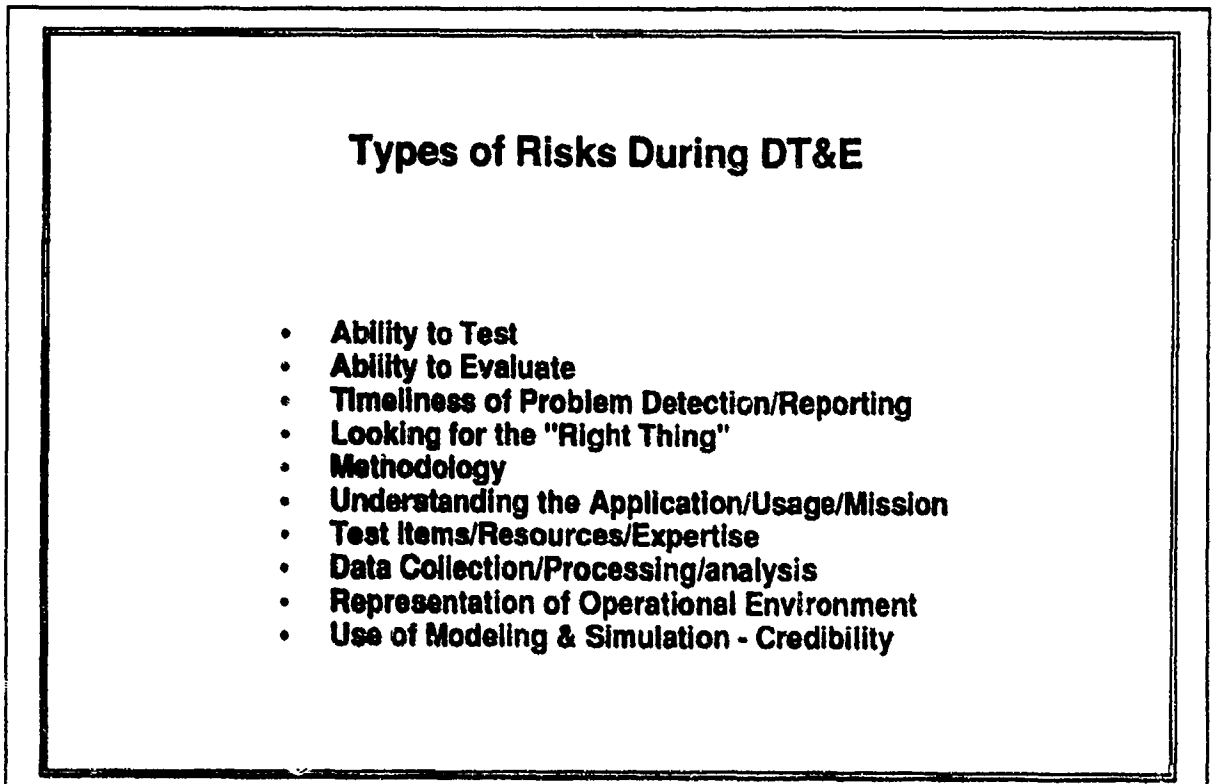


FIGURE 4

- Combined Arms Emphasis
- Fewer New Starts/More P3I
- Acquisition Corps (Open doors to funds for education and training)
- Changing Doctrine
- DT Certification for OT

CONSTRAINTS:

- Reduced Funding
- Environmental
- Reduced Time
- Congressional Oversight
- Expanded Envelopes (e.g., safety on lasers)
- Uncertainty of Threat
- Combined Arms Emphasis
- More P3I — Modifications
- Increased Complexity/Software Emphasis

STEP 5

Question: In light of the "external conditioners", what do we see as the strengths and weaknesses of the evaluators in DT&E to respond?

Further brainstorming identified the following "internal conditioners" as items over which the T&E community has some "control." Some of these were identified as "strengths" those items which are beneficial in the DT&E evaluator community today, towards reducing risks, and upon which we could capitalize for further reducing risks. Others are identified as "weaknesses" because they are limiting the ability of evaluators to reduce risks, and are viewed as necessary to correct, overcome, or improve.

INTERNAL CONDITIONERS (WITHIN CONTROL OF THE T&E COMMUNITY)

STRENGTHS:

- Advanced Test Technology
- Advanced Evaluation Technology
- Multi-Service Initiatives
- Modeling and Simulation Validation, Verification, and Accreditation Process
- Defense T&E Professional Institute (DTEPI)
- Test and Evaluation Community Network (TECNET)

WEAKNESSES:

- Poor Appreciation by Decisionmakers
- Modeling and Simulation Credibility (not far enough along)
- Doctrine Testing and Evaluation
- Infrastructure Decay
- Joint Program — Service Funding
- Lack of Independence of Evaluator from PEO/PM
- DT&E Evaluator/User Relationship
- Limited Test Facilities/Ranges/Expertise
- Adequacy of DT&E Under "Operational Conditions"
- Integration vs. Component
- PEO/PM Minimal Support of DT&E
- Credibility: Conflict of Interest Perception
- Lack of "Validated" Threats: Models, Simulations, Simulators
- Interoperability of Modeling and Simulation/Standard Inputs/Outputs

- Lack of Models: System or Engineering Level
- Inadequate Documentation on Modeis
- Lack of Test Programs to Validate Models (Model-Test-Model)
- Access to Emerging and Past Data/Known Format
- Smarter Evaluation Techniques to Compensate for Test Equipment Limitations
- Need for Evaluation Plan to Drive Test Plan
- Lack of Standardized Procedures (Range Tests)
- Education/Training Needs
- Intern Programs (Accelerate On-The-Job Training)
- Funding for Early Involvement
- Access to Proprietary Information
- Inadequate Knowledge of Intended Use of System

STEP 6

Question: Given the external and internal conditioners, and an unlimited opportunity to change the evaluator's role, resources, tools, expertise, and/or image to reduce risks during DT&E, how would the evaluator like to be perceived in five years? In other words, what would our "vision" be for the evaluator in five years, if we could dream?.

Breaking up in small groups of three or four, each group developed proposed elements of a "vision." Each group described their elements to the entire Working Group III until the final elements were unanimous among all participants, and remaining elements discarded. Figure 5 is the resulting vision.

STEP 7

Question: What events, actions, decisions, or resources would have to happen (recommendations) in order for us to realize our vision?

With participation of the entire Working Group III, each element of the vision was reviewed until unanimity was obtained on each action required. In order to achieve the vision to reduce the inherent risks during development and OT&E through the evaluator, the following strategic actions should be accomplished:

- Establish credibility — be known as the "honest broker."
- Develop early DT&E and OT&E evaluation concepts (to precede the Test and Evaluation Master Plan (TEMP)).
- Perform continuous evaluation.
- Provide education and training for evaluators.

VISION

IN FIVE YEARS THE INHERENT RISKS DURING DT&E AND OT&E WILL HAVE BEEN REDUCED DUE TO DT&E EVALUATORS:

- **HAVING INCREASED EVALUATOR CREDIBILITY**
- **DEVELOPING EARLY DT&E AND OT&E EVALUATION CONCEPTS PRIOR TO TEMP**
- **PERFORMING CONTINUOUS EVALUATION**
- **HAVING INCREASED EDUCATION AND TRAINING**
- **DRIVING IMPROVED TEST RESOURCES/FACILITIES**
- **HAVING CONSISTENT T&E PROCESSES AND PRINCIPLES ACROSS SERVICES**
- **HAVING IMPROVED ACQUISITION COMMUNITY RELATIONSHIPS**
- **HAVING CLOSER NETWORKING WITHIN THE T&E COMMUNITY**

FIGURE 9

- Improve test resources and facilities.
- Have consistent test and evaluation processes and principles across the military services.
- Improve organizational relationships among developer, tester, user, and evaluator.
- Provide closer networking within the T&E community.

RECOMMENDATIONS

From the vision and strategic actions above, Working Group III recommends the following:

CREDIBILITY

The DT&E evaluator must establish credibility so that he can be looked upon as an "honest broker." This is crucial to achieving the "vision."

RECOMMENDATION 1: Satisfy the customer with what he needs.

The evaluator must recognize that the customer doesn't always know what he wants, nor can he define or describe it, but he recognizes it when he sees it. Whether the customer is a program manager, user, OSD, a contractor, an operational test agency, Congress, or whoever, the "evaluator" must be relied on to identify the

really critical issues and to present the results of his evaluation in meaningful and understandable ways. This suggests that even though he may use highly sophisticated analytic techniques, he must develop skills to communicate (written and orally) in the terms and from the perspective of the customer.

RECOMMENDATION 2: Assist the user in developing evaluatable requirements.

Traditionally, the question has been whether a requirement "is testable?" Recognizing that as systems become more complex and software intensive, and that as capabilities to perform tests become more limited or can't be performed because of funding, environmental or other constraints, many of the critical issues will only be answerable through analysis. If the evaluator has an adequate toolbox of analysis tools, he will be able to address many issues that can not be tested, but are "evaluatable." The evaluator should participate with the user (maybe eventually replacing the tester) in defining his requirements.

RECOMMENDATION 3: Get acceptance of modeling and simulation (M&S).

Modeling and simulation (M&S) could be one of the most important tools for evaluating testing-constrained systems, but this approach still lacks widespread acceptance that the results can be believable or credible (particularly among congressional staffers). Establishing the credibility of M&S through demonstrating comparability or correlation of results with test programs should be an early emphasis in the T&E community.

RECOMMENDATION 4: Establish exit criteria reflecting the operational requirements document (ORD).

DT&E evaluators have tended to focus on technical requirements, usually as required by the specifications of contracts. Issues of effectiveness and suitability are seldom, if ever, required in specifications; however, it is in these issues that the user and most decision-makers are interested. How a new system is to be deployed is reflected in the ORD, and it will be this document that will form the basis of the OT&E. Contractor and military evaluators during DT&E need to include these issues in their evaluations as exit criteria to certify that the system is ready for OT&E. This will enhance the image of the evaluator (his credibility) of truly looking at all aspects of the systems being considered for fielding.

RECOMMENDATION 5: Develop realistic measures of performance (MOPs), measures of effectiveness (MOEs), and thresholds.

DT&E evaluators need to consider the expected state of maturation of a system or component when establishing what the MOPs and MOEs are to be, and should specify thresholds accordingly. While maintaining their independence, they should also be helping with the development. Program managers, for instance, should want to hear what the evaluator has to say — a feeling should be nurtured that the evaluators' criteria are fair to the degree of maturity.

RECOMMENDATION 6: Coordinate MOPs and MOEs with users.

It is still the general case that the DT&E evaluator has no contact with the

user. It would help the credibility of the DT&E evaluator if the MOPs and MOEs he develops are coordinated with the user in order to ensure that his concerns are being addressed.

RECOMMENDATION 7: Assist in solving problems (vice only identification).

The DT&E evaluator should consider himself as a part of the "development" team, and not only identify problems, but help in solving them. He is in a unique position of having insights into performance problems having isolated them, he can most likely see possible solutions and evaluate the alternatives. Again, this can help his evaluation services to be sought.

RECOMMENDATION 8: Improve customer relationships.

As in the total quality initiatives of the Department of Defense and the commercial sector, the DT&E evaluators need to work to improve their relationships with their customers. Customers may include the testers, users, program managers, contractors, the respective service and/or OSD, and congress; but, ultimately, it is the soldier, sailor, airman, or marine who will be risking his life in the system to which the evaluator should be serving. The more that evaluator relates and coordinates with his customers, and responds to his needs, the more credibility he will create. He can do this without sacrificing "independence."

EARLY DT&E AND OT&E CONCEPTS

The thrust of this element of the "vision" is for the evaluator to develop an evaluation concept, strategy, or plan as one of the earliest steps, and this would be used as a basis for a test program to be included in the TEMP. In essence, the evaluation concept should drive whatever testing, modeling and simulation, experiments, physical modeling, and analyses that should be accomplished throughout the entire development program. This evaluation concept should be based on a "design of experiments" type of approach.

RECOMMENDATION 9: Institutionalize the requirement to develop an early "evaluation concept" through policy and regulations.

Development of an early evaluation concept will require an early expenditure of resources and management emphasis. To be effective, that is, in order to drive the requirements in the TEMP, this effort will likely precede the availability of program funding, and would have to come out of overhead funding. Additionally, this effort would have to compete with on-going evaluations usually having high priority. Therefore, the requirement must included in policies and regulations among the services.

RECOMMENDATION 10: Develop a contents outline for the early evaluation concept.

A recommended outline of contents of what should be contained in early evaluation concept should be developed by members of both the DT&E and OT&E communities, military and contractors. Perhaps as a subject for a future ITEA-MORS workshop, this outline should encompass results from a "design of experiments" type approach, and include a plan for the sources of all data to be used the evaluation; eg, modeling and simulation, experiments, testing, exercises, combat,

prototypes, previous tests (OT and DT), existing databases, and intelligence.

RECOMMENDATION 11: Utilize an evolutionary approach involving phases in the early evaluation concept.

The early evaluation concept should recognize that the system will evolve over time, and that the final expected results are not likely to be seen during the early phases of the program. The concept, then, should allow for evolving maturity of components as well as the system.

RECOMMENDATION 12: Develop balance between system as well component evaluation.

Many programs develop or modify systems (components) that are to become a part of a larger system. The early evaluation concept must contain provisions for evaluating the "component" as a part of the larger system: ie, system compatibility, software interoperability and protocols, interference, increases to operational burden, changes to visual and electro/optical signatures, operational security impacts, overall system performance changes, etc.. A balance between evaluating the component and the system into which it is to be integrated should be included in the evaluation concept.

RECOMMENDATION 13: Include the usage of modeling and simulation (M&S) in the evaluation concept.

Acquiring data from M&S should be a cost effective alternative to actual testing when the cost or availability/constraints of test facilities/ranges will preclude getting sufficient sample sizes for a meaningful evaluation. The use of M&S could afford the capability of helping in the "design of experiments" and to determine which data points that will be needed in actual testing, and to help verify the model with actual data. Early agreement on using the "model-test-model" approach will be essential when models will have to be developed.

CONTINUOUS EVALUATION

In the "vision" evaluation will be continuous throughout the life of a system (a weapon) — it will not stop when the basic development or acquisition has been accomplished. The evaluators will continue to evaluate the system to determine its viability in light of new or changing threats, new doctrines and tactics, technology breakthroughs or transfers to potential opponents, political and social pressures, security compromises, or for possible new applications or missions. This suggests that "continuous evaluation" become institutionalized, and that funding/resources be dedicated for this purpose.

RECOMMENDATION 14: Institutionalize "continuous evaluation" for the life of all military systems.

The military services should adopt policies and fund for resources to provide continuous evaluation. Periodic continuous evaluation reports should be required.

RECOMMENDATION 15: Built-in test equipment and data recording should be incorporated into the design of all military systems.

Built-in status and fault isolation capabilities along with a data bus are becoming more commonplace in designs; an

extension of these capabilities could provide a means of providing data for conducting continuous evaluation. Providing sensors and ports for data required by the evaluator, and a means for recording could collect data on the system during actual usage under a variety of conditions.

RECOMMENDATION 16: Share development and operational data among development and operational evaluators.

Data collected during development and operational tests and modeling and simulation should be made available to all evaluators in order to supplement data they have obtained from their respective test and M&S programs. Extra copies of data dictionaries and evaluation reports should be retained for providing to counterpart evaluators.

RECOMMENDATION 17: Make data available to evaluators from all sources: modeling and simulation, testing, exercises, training, and combat.

Data collected from any and all sources will be useful in performing continuous evaluation. Any data collected during actual combat would be the most useful, especially if it can be annotated with information on the enemy, allies, intelligence, and the combined arms involved in the operation.

RECOMMENDATION 18: Retain databases for the life of the system for use in continuous evaluation.

Databases on all aspects of the system should be archived and made accessible to evaluators at any time required during the system life. This suggests that a management system will need to be created, and storage facilities provided. However, access to this data will aid in performing continuous evaluation.

EDUCATION AND TRAINING FOR EVALUATORS

In order to realize the "vision," evaluators must have access to education and training that provides the foundations for professionalism in evaluation. Evaluation requires some specialized skills and knowledge that will differentiate the evaluator from a tester.

RECOMMENDATION 19: Develop an evaluator's syllabus for providing to the Defense T&E Professional Institute (DTEPI).

Evaluators from the military services and industry need to collectively develop a syllabus of the types of courses and experience that will provide the necessary tools, skills, and knowledge for evaluators to have. (Possibly a topic for an ITEA-MORS workshop.) An opportunity exists that this syllabus could be used as a basis for establishing requirements for the Acquisition Corps. DTEPI (provides support to the Dept of Defense and its contractors) can be requested to identify sources for the courses and training required, or to set up the mechanism for creating the courses.

RECOMMENDATION 20: Develop a test and evaluation degree.

Support should be given to DTEPI in their efforts to set up a degree program in test and evaluation. An opportunity should

be afforded to students to get the specialty courses to specialize in evaluation.

RECOMMENDATION 21: Provide specialized training in operations research, systems analysis, and system engineering for evaluators.

A syllabus for evaluators must as a minimum provide the elements of operations research, systems analysis, and system engineering. An emphasis of these subjects should be design of experiments. An introduction to applications of chaos theory, Bayesian techniques for minimum sample sizes, dealing with "messy data," and modeling and simulation should also be included.

RECOMMENDATION 22: Provide operational exposure for evaluators.

As a part of their training, evaluators should spend some minimum time in the environment with the soldiers, sailors, airmen, or marines that will be using the types of systems that they will be evaluating. Each of the military services should have a program that permit this type of exposure to contractor and military service evaluators.

RECOMMENDATION 23: Implement an evaluator's intern program.

Each of the military services and contractors should establish intern programs for evaluators that would recruit new college graduates, and place them with evaluators. This would give them opportunities to learn from the experts, while at the same time provide some data analysis under continuous evaluation.

RECOMMENDATION 24: Educate decision makers on the evaluator's tools, and their uses and limitations.

A unit on what evaluators can tell decision makers should be included in management training and education. This unit should also stress, more importantly, what the limitations of the evaluator are so that the decision makers don't have unrealistic expectations of what the evaluator can provide, and how to interpret the information. This unit should be included in all Defense Systems Management College (DSMC) programs, as well as each of the systems acquisition and management courses of each of the military services. This unit could be designed as a topic at a future ITEA-MORS workshop.

RECOMMENDATION 25: Develop training for evaluators on the risks of manufacturing processes.

The new DoD Instruction 5000.2 requires that risks assessments be performed on new manufacturing processes, a new area for evaluators. The requirements for such a course should be passed to DTEPI to find sources for such training. The actual requirement should be developed as a topic for a future ITEA-MORS workshop.

IMPROVED TEST RESOURCES AND FACILITIES

Key to the vision is having adequate test resources and facilities to obtain critical data points for evaluation. No matter how much progress is made in modeling and simulation, and other forms of analysis, without being able to test some of the points (eg, as in model-test-model), confidence in the results, findings, and conclusions of the

evaluator will be (and should be) suspect. Working Group III unanimously recognized the inadequacies of today's test facilities to acquire data points that are essential to performing high confidence evaluations.

RECOMMENDATION 26: Link evaluation requirements to investments in new test capabilities.

In determining the requirements and priorities of new test facility investments (eg, ranges, threat simulators, targets, anechoic chambers, wind tunnels, signature measurement facilities, electromagnetic pulse simulators), evaluation requirements should be a primary consideration; particularly in light of Project Reliance, where the military services will increasingly rely on each other for test capabilities rather than duplicate them, and testing to collect data will be done at a number of different locations. A single evaluator for any given program may have any or all Department of Defense, other government (eg, NASA, Department of Energy), or industry test facilities available to provide data for his evaluation. Further, in conjunction with using tools such as modeling and simulation, the evaluator may not require the use of certain test facilities to the extent as he has previously. Using the recommended approach of developing an "early DT&E and OT&E evaluation concept," as also recommended by Working Group III above, based on "design of experiment" type of test optimization approach, investments should be based on the most critical evaluation needs for test data to reduce risks.

RECOMMENDATION 27: Tie test facility investment requirements to technologies vice programs.

Test investment requirements are usually driven by needs to test a specific program, but, because of the long lead time to develop, install, and calibrate most test facilities, they are usually too late to provide data at early stages of the program when chances to reduce risk are greatest. Knowledge of the types of technologies that will have to be evaluated should be used as a basis for identifying and prioritizing test facility investments.

RECOMMENDATION 28: Advocate for the Office of the Secretary of Defense (OSD) to fund for technology edge and common requirements.

The military services seldom fund for test facilities that are not tied to specific programs (for example, for new technologies or to develop a new test technology that will require a long lead-time), or if another service has requirements for the same capability (they will prioritize it low in hope that the other service will fund it). Justifications for such facilities have to compete for funds against justifications for new weapons and /or people programs. For technology requirements, for new test technologies, and when more than one military service requires a capability, OSD should advocate and manage separate appropriations to fund these types of requirements based on evaluation needs.

RECOMMENDATION 29: Investments in test facilities should emphasize operational conditions/environment.

Test facilities to perform DT&E should be designed/developed such that they are representative of the operational conditions and environment for which systems will be evaluated to certify that they

are ready for OT&E. To avoid surprises during OT&E, the DT&E evaluator should use test facilities from which to collect data that will give him the most insight into how the system is expected to perform when deployed. While few contract specifications will specify requirements for effectiveness or suitability, contractors and DT&E evaluators must anticipate that a system must demonstrate its ability to meet the performance requirements under operational conditions against operational issues.

RECOMMENDATION 30: Provide mobile test facilities

The best return on investment for new test facilities is to make them so that they can be easily relocated to be used in the most operationally representative locations. When possible, test facilities should be designed/developed such that they have a standard interface with various locations, and can be interoperable with other facilities. This will afford the evaluator the opportunity to acquire data from tests that are most representative of the operational environment.

RECOMMENDATION 31: Make models and simulations interoperable.

To provide the best return on investment in the development and use of models and simulations (M&S), they should be reusable for a large number of applications, and be usable in conjunction with other simulations of higher and lower levels. The initiative by OSD to develop a common modeling and simulation architecture in the Joint Modeling and Simulation System (J-MASS) program should be supported be the military services and industry in order to establish standard M&S input and output characteristics that will permit system level models (of friendly and threat systems) to be used in platform level simulations, platform level in mission level, and mission level in theater or campaign level simulations. This hierarchy of simulations using reusable models in place of redeveloping new ones for use in different simulations can save estimated hundreds of millions dollars being spent on duplicating models, and provide consistency in results through use of validated models.

RECOMMENDATION 32: Obtain system models as early contract deliverables.

System level models of proposed new systems or improvements should be developed as one of the first products or deliverables from any contractual development program (ideally as part of the proposal), and be evaluated like any other test item. This will permit early evaluation in order to lower risks in the development. Delivered in accordance with standard architecture formats (such as being developed under the Joint M&S System (J-MASS) program) will permit these models to be evaluated in the context of higher level (platform, mission, theater) simulations to predict military worth or contributions the system or improvement will make to combat scenarios. These system models should also updated throughout the life cycle of the system or improvement.

RECOMMENDATION 33: Provide facilities for running simulations.

Investments should be made in a few world class facilities specifically designed for the use by evaluators in running higher level simulations to include totally software, man-in-the-loop, and/or system- or

improvement-in-the-loop. To be available for use by contractor and military service evaluators, these facilities should provide for high powered evaluations before embarking on new phases (Milestones) of development.

RECOMMENDATION 34: Develop standardized access to databases.

A standard for databases for archiving of data from all test and evaluations and all other sources needs to be established for use by contractors and the military services. This standard is to provide a consistent means of accessing data in order that evaluators aren't stymied or impeded in bringing all available data to bear in their evaluations. Development of this standard could be a topic for a future ITEA-MORS workshop.

RECOMMENDATION 35: Establish a library of accredited models.

Models that prove to be useful, that are verified and validated, and that have been accredited for specific uses should be maintained in a library for use by contractor and military service evaluators. Updated with actual data, and maintained under configuration control procedures, a library system, such as available under SURVIAC, would serve as a clearinghouse for all models and simulations used in evaluations.

RECOMMENDATION 36: Develop a real-time analysis capability throughout the life cycle of each program.

As a part of every development program (as part of the cost), capability for real-time analysis should be provided to permit its continuous evaluation throughout its entire life, starting with concept tradeoffs, continuing through the design phase, DT&E, OT&E, deployment, training, and combat mission planning, and post-mission analysis. This capability should consist of all models and simulations, databases, and look up tables appropriate to analysis of the system or improvement.

CONSISTENT TEST AND EVALUATION PROCESSES AND PRINCIPLES

Concerns were expressed by participants in Working Group III that each of the military services have different standards, procedures, methodologies, practices, and processes in test and evaluation. What might be acceptable within one military service (eg, participation in T&E by contractors involved in developing the system or improvement) will not be acceptable by another military service. A common plea for consistency was made.

RECOMMENDATION 37: Standardize T&E processes and principles across the military services.

Support efforts by the military services through initiatives by such groups as the Joint Commanders Group for Test and Evaluation (JCG(T&E)) of the Joint Logistics Commanders, the Multi-Service Test Investments Review Committee (MSTIRC), the Test and Evaluation Community Network (TECNET), the Defense T&E Professional Institute (DTEPI), and the Range Commanders Council (RCC). Also, joint programs such as the Joint Modeling and Simulation System (JMASS) program and Project Reliance are helping to "level" the differences in the T&E community between the military services.

OSD has a number of initiatives such as the Defense Test and Evaluation Strategy and the Test Resource Master Plan (TRMP), as well as a number of programs to standardize test methodologies. Building on these efforts, and initiating new ones, with a focus on evaluation, analysis, and modeling and simulation, can contribute to standardizing policies and procedures.

IMPROVED ORGANIZATIONAL RELATIONSHIPS

Total quality management philosophy and initiatives within the Department of Defense and industry stress satisfaction of the "customer" and demanding quality from "suppliers." To improve the effectiveness of the evaluator to reduce risks, these same principles apply.

RECOMMENDATION 38: Initiate actions to improve the organizational relationships among the developer, tester, and user with the evaluator.

A goal of the DT&E evaluator should be to help the development process (while maintaining his "independence"). He should also be a watchdog to ensure the needs of the user will be met. At the same time, he should be demanding in the quality of data supplied by the testers no matter who provides it (industry, various military and government agencies). The testers have made strides to improve their relationships with developers and users whereas the evaluators have yet to see much improvement. The evaluators should become proactive to seek what the developers and users expect (would like) from them, and establish metrics (eg, feedback) on how well they are doing. Likewise, evaluators should institutionalize methods for providing feedback to testers. Determining specific actions could be a topic for a future MORS-ITEA workshop.

CLOSER NETWORKING WITHIN THE T&E COMMUNITY

In spite of a number of initiatives in the past to improve communications among testers and evaluators, there was a general feeling among Working Group III that it still doesn't exist. Sharing of information within the T&E community about testing and evaluations, techniques and methodologies, test facilities, education and training opportunities, and policies and accepted practices is considered essential for the evaluators to contribute to risk reduction.

RECOMMENDATION 39: Use of the Test and Evaluation Community Network (TECNET) should be promoted.

TECNET is available to all Department of Defense personnel and industry personnel having active defense contracts. Having both unclassified and classified modes, and several databases and bulletin boards for a broad spectrum of information pertaining to T&E (eg, existing and proposed test facilities and resources, policies and guidance, expertise database, and job opportunities) as well DoD-wide electronic mail and fax capabilities, TECNET should be used by all evaluators.

RECOMMENDATION 40: Capitalize on the interService agreements from Project Reliance to open communications with evaluators of other military services working in similar areas.

An unprecedented channel of communications is being afforded by Project

Reliance. For the next several years, T&E personnel from all military services will be working together to determine how to minimize the number of test facilities required by relying on each other. Evaluators should participate in these activities, and determine opportunities to rely on each other for certain expertise, skills, tools, and methodologies.

CONCLUSIONS

Improvements in evaluation during DT&E could reduce risks inherent in system development and in having unfortunate surprises during OT&E. Working Group III has developed a number of recommendations to make evaluation more effective in reducing risks. All of these recommendations are considered reasonable and doable within current budget constraints, and are consistent with external events and policies. They are viewed as an opportunity that, if embraced by the Department of Defense and Industry, our vision could be realized in five years.

WORKING GROUP III PARTICIPANTS:

NAME	ORGANIZATION
Lt. Col. Joseph F. Auletta	Foreign Technology Division
Mr. Michael D. Bates	OAS(AFSC)
Major David H. Berg	HQ SAC/XPSW
Mr. Irvin E. Boyles	US Army OPTEC
Mr. Gary Bridgewater	SAIC
Mr. Robert E. Dufresne	Naval Underwater Systems Center
Mr. Terrence J. Erickson	Alliant Tech Systems, Inc.
COL Bernard B. Ferguson	OSD(DDR&E(T&E))
Lt. Col. Richard T. Frank	HQ AFOTEC/XPX
Col. Frederick L. Green	HQ AFOTEC/TEZ
Mr. Roger W. Johnson	University of Central Florida
Mr. John A. Kelley, Jr.	Raytheon
Mr. Benjamin F. King	US AMSAA
Mr. James E. Leach	Lockheed Sanders
Mr. Robert K. Lebowitz	MAR, Inc.
Mr. Richard R, Ledesma	OSD(DDR&E(T&E))
Mr. Vincent P. Merz	VEDA, Inc.
Mr. Sidney Polk	The MITRE Corporation
Mr. Matthew T. Reynolds	ms Command
Mr. Bruce F. Rogers	SAIC
Mr. Jerome C. Rosenberger	Institute for Defense Analyses
Mr. Marshall J. Tino	Naval Surface Warfare Center
Mr. Larry L. West	US Army TECOM
CAPT Douglas P. Williams	Department of the Navy
Mr. James E. Youngblood	US AMSAA
Mr. Charles W. Zang	BDM International, Inc.

WORKING GROUP IV

EVALUATION TECHNIQUES TO OVERCOME THE LIMITATIONS IN SUITABILITY TESTING

Co-chairs: Dr. John Foulkes
Test and Evaluation Management Agency
HQ Department of the Army

Mr. Dick Gunkel
SIGMA PLUS, Incorporated

Rapporteur: Mr. Gene Hauze
MITRE Corporation

EXECUTIVE SUMMARY

The purpose of Working Group IV was to improve the use of evaluation techniques relative to suitability testing of military systems by examining specific evaluation mechanisms to overcome limitations in developmental and operational suitability testing. The group's objectives were to establish a community within both the government and industry for the promulgation of suitability evaluation processes. This in turn would establish credibility of the processes with the decision makers. It would also provide a baseline for developing the mechanism to institutionalize suitability evaluations.

The working group set out to identify the problems, concerns, and issues that relate to suitability evaluations. As a result, forty issues were identified by the group. Of those, nine key issues were identified as requiring the most urgent attention. These issues were: 1) The need to identify an Office of the Secretary of Defense (OSD) focal point for the suitability evaluation process; 2) The need to increase the impact of suitability evaluations; 3) The need to determine the parts of suitability evaluation that can be accomplished or improved by modeling and simulation; 4) The need to improve the ability to measure the elements of and to evaluate operational availability; 5) The need to standardize definitions, data collection and database design (particularly associated with Reliability, Availability, and Maintainability (RAM); 6) The need to improve the ability to evaluate software intensive systems with a particular focus on software RAM; 7) The need to improve guidance on the application of reliability growth (positive or negative); 8) The need to have consistent guidance on the reporting of RAM factors and measures (e.g., point estimates versus confidence intervals); and 9) The need to clarify guidance for

evaluating suitability where contractor support is planned.

Recommendations were developed to address each of the nine issues stated above. Interestingly, these recommendations basically fall into the single category of a need for more management attention and structure to get the suitability evaluation job done. This was further supported by numerous examples of excellent suitability evaluations. The capability exists, so we now need to find a way to apply it consistently and effectively. There is a particular need to encourage suitability evaluation in developmental testing. The payoff and leverage for identifying and fixing suitability problems early has been documented at well over a 10-to-1 return on investment ratio.

PURPOSE

The general purpose of Working Group IV was to examine the process of suitability evaluation of military systems and to address the existing limitations relative to developmental and operational suitability testing. The group consisted of the two co-chairmen and 20 participants. The group examined specific processes, tools, mechanisms, and procedures as well as specific examples that can be or have been applied such as modeling and simulation, comparability analysis, data structures, reliability growth methodologies, and other data analysis techniques.

The group was asked first to accept the premise that there is a problem — suitability evaluation is under utilized and viewed by many as a non sequitur, yet is becoming more important and critical due to increasing costs, decreasing budgets, and changes in design, development, and manufacturing processes; secondly, solutions exist or can be developed. Operating under this premise, we would be able to: a) provide a forum that clearly establishes evaluation as a proven and valuable suitability tool; b) provide clear documented examples; and c) provide an outline of how to plan, develop and apply evaluation tools, thereby increasing the application of evaluation in this critical area and thus greatly improve the suitability process.

The basic objectives of the working group were to establish a community within both the government and industry to promulgate suitability evaluations and to provide clear, concise, and convincing evidence that proves the need and provides the evidence in a form that can be used to establish credibility with the decision makers who will be the users of the results. The group was also asked to provide and initial baseline for developing and refining the tools, techniques, and processes to institutionalize suitability evaluations.

ISSUES IDENTIFIED

The group accomplished this objective after two days of intense discussions by establishing the following forty issues:

1. There is a need to develop a generic suitability evaluation model that could be tailored by users to evaluate specific systems.

2. There is a need to better define early user requirements.

3. There is a need to derive a test and evaluation methodology to address component failures associated with cumulative effects.

4. There is a need to increase attention given to the suitability elements of the logistics support system.

5. There is a need to improve commonality in defining measures of effectiveness (MOEs).

6. There is a need to improve comprehensiveness in Operational Testing (OT) Objectives and Reports regarding operational suitability.

7. There is a need to change suitability measurements reporting to include more than simple point estimates.

8. There is a need to more effectively utilize small sample sizes in decision making.

9. There is a need to more effectively integrate OT data with ground/simulation data.

10. There is a need to standardize models and data collection methodology/techniques.

11. There is a need to close the gap between mission profile requirements and OT requirements.

12. There is a need to weight reliability measurements to mission essential functions.

13. There is a need to improve the ability to measure the elements of and to evaluate operational availability.

14. There is a need to develop metrics for measuring the logistics burden associated with attaining a required operation availability.

15. There is a need to improve guidance on the application of reliability growth (positive or negative).

16. There is a need to integrate damaged parts estimates derived from modeling into logistic support requirements.

17. There is a need to lessen the operational effectiveness focus of operational testing.

18. There is a need to determine a methodology to demonstrate high reliability requirements given limited resources. There is a need to determine whether development test data can be utilized to achieve this end.

19. There is a need for consistent guidance on the reporting of RAM factors and measures (e.g., point estimates versus confidence intervals).

20. There is a need to increase the impact of suitability evaluations. There is a need to determine whether this can be accomplished via cost effectiveness modeling.

21. There is a need to improve the ability to evaluate software intensive

systems with a particular focus on software RAM.

22. There is a need to improve the operational suitability evaluation of non-developmental items (NDI).

23. **There is a need to clarify guidance for evaluating operational suitability where contractor support is planned.**

24. There is a need to address the inflation of mean time between failure (MTBF) estimates by contractors.

25. There is a need to provide better access to logistic support analysis data.

26. There is a need to increase the utilization of existing suitability evaluation models.

27. There is a need to tie operational suitability to operational effectiveness.

28. There is a need to better utilize historic suitability data.

29. **There is a need to determine the parts of suitability evaluation that can be accomplished or improved by modeling and simulation.**

30. (Combined with #29.) There is a need to identify the availability, selection and use of suitability models.

31. There is a need to include cost in the evaluation of suitability. There is a need to identify the responsible agent(s) for accomplishing this effort.

32. There is a need to derive a methodology for including suitability evaluation requirements in test planning.

33. There is a need to increase commonality among services with respect to suitability definitions.

34. **There is a need to standardize definitions, data collection and database design (particularly with RAM).**

35. There is a need to provide greater attention to correcting MANPRINT problems found during testing.

36. There is a need to increase the utilization of lessons learned from modeling and simulation to support testing.

37. There is a need to increase the use of Early User Test and Evaluation (EUT&E) and Feasibility Demonstration Test and Evaluation (FDT&E) to support early evaluations of operational suitability.

38. There is a need to breakdown the data sharing barrier between development testing (DT) and OT.

39. There is a need to develop a methodology/technique for utilizing Test, Measurement, and Diagnostic Equipment (TMDE) data.

40. **There is a need to identify an OSD focal point for the suitability evaluation process.**

CRITICAL ISSUES AND RECOMMENDATIONS

Of the above forty issues, the group identified nine key issues (those highlighted above in bold) and developed a recommended solution for each.

CRITICAL ISSUE: There is a need to identify an OSD focal point for the suitability evaluation process. Even though the group could not come to a consensus on a recommended solution for this issue, it is recognized that despite recent emphasis on the suitability evaluation and process, operational effectiveness is still receiving the majority of the attention as far as upper level management champions are concerned. The solution most favored recommended the creation of positions in DDDR&E(T&E) and DOT&E to take the lead in planning and overseeing the execution of the process.

CRITICAL ISSUE: There is a need to increase the impact of suitability evaluations. There is a need to determine whether this can be accomplished via cost effectiveness modeling or not. The group recommended, where possible, that suitability issues be presented in terms of performance and cost. The evaluators must present suitability, not in isolation, but in terms of performance and cost (e.g., bring the Cost and Operational Effectiveness Analysis (COEA) into the action).

CRITICAL ISSUE: There is a need to determine the parts of suitability evaluations that can be accomplished or improved by modeling and simulation. The group recommended the community utilize and expand upon existing DOT&E guidance on operational suitability, and emphasize early implementation of this guidance in the acquisition process. The suitability process must also participate in the Defense Modeling and Simulation Office's (DMSO) activities regarding funding and support of modeling and simulation technologies.

CRITICAL ISSUE: There is a need to improve the elements of and to evaluate operational availability. What elements can be measured in testing, and how do we address the missing elements? What are the critical elements? For example, one factor which has a significant impact on availability for most new systems is integrated diagnostics (ID). How we can improve ID evaluations. The group advocated the establishment of a working group to develop a consistent methodology.

CRITICAL ISSUE: There is a need to standardize definitions, data collection and database design (particularly associated with RAM). The group could not come to a consensus as to how to resolve this issue. However, it was decided that an executive level working group could be established to develop a phased plan requiring standardization of definitions, data collection, and database design. This would entail a "bottom up" approach, that is, beginning in each of the Services, and then expand to include the Department of Defense (DOD) and then industry.

CRITICAL ISSUE: There is a need to improve the ability to evaluate software intensive systems with a particular focus on software RAM. Our acquisition systems are becoming more and more software intensive.

It was acknowledged that much is going on within DOD and the Services in this area already. However, it was felt that a significant and immediate acceleration in DOD-wide emphasis, coordination, and incorporation of evaluation methodologies and metrics for evaluating software must be done.

CRITICAL ISSUE: There is a need to improve guidance on the application of reliability growth (positive and negative). It was recommended that service centers of excellence be tasked to identify and refine existing service methodology (e.g., update MIL HDBK 189, Reliability Growth Management).

CRITICAL ISSUE: There is a need for consistent guidance on the reporting of RAM factors and measures (e.g., point estimates versus confidence intervals). The difference in reporting of RAM measures was a major concern to the group. The majority of the participants recommended that a joint service study be commissioned to investigate whether RAM point estimates versus confidence levels impact decision making.

CRITICAL ISSUE: There is a need to clarify guidance for evaluating suitability where contractor support is planned. What are the rules and procedures, and should contractor support performance be evaluated along the same lines as government support performance? The recommendation was to review service documentation to assure it provides guidance on evaluating contractor support.

SUMMARY

It isn't too difficult to see that the above recommendations basically fall into the category of increasing management attention and creating sufficient structures to get the suitability evaluation job done. The creation of an OSD-level cell to serve as the manager and to foster the process would go far in solving most of our suitability evaluation shortfalls.

As a result of conducting this working group, it is hoped that the participants took away with them a firm belief in becoming active in the pursuit of improving suitability evaluations, and that each member will aggressively pursue improving the application of suitability evaluation techniques. We believe that sufficient information was acquired during the two days to convince the test and evaluation community and the decisionmakers of the existence of many significant issues relative to suitability evaluation, and also the need to implement the recommendations to improve suitability evaluations across all phases of the acquisition process.

WORKING GROUP IV PARTICIPANTS:

NAME	ORGANIZATION
Mr. Frank J. Apicella	US Army OPTEC
Dr. Stanley F. Bolin	US Army OPTEC
Mr. Michael S. Bridgman	Logistics Management Institute
Mr. George Chernowitz	American Power Jet Co.
Professor John D. Claxton	Defense Systems Management College
Mr. James B. Duff	COMOPTEVFOR
Dr. Paul M. Ellner	US AMSAA
Dr. John B. Foulkes	HQDA/TEMA
Dr. Arthur Fries	Institute for Defense Analyses
Mr. Eric Grape	US AMSAA
Mr. Richard A. Gunkel	Sigma Plus, Inc.
Mr. Gene E Hauze	The MITRE Corporation
Major Johnny R. Hays	HQ SAC/XPSW
Ms. Joyce A. Hires	US Army TECOM
Dr. Dong S. Kim	US Army CAA
MAJ Michael J. Leahey	MCOTEA
Mr. Thomas R. Metz	OSD(DDR&E(T&E))
Mr. David E. Mortin	US AMSAA
Mr. David A. Oppenheimer	RA Corporation
Mr. Thomas J. Parry	OASD(P&L)/L/WSIG
Lt. Col. Stalker E. Reed	HQ AFOTEC/TEZ
Capt. James W. Wisnowski	HQ AFOTEC/LG4W

WORKING GROUP V

EVALUATION AS A CRITICAL ELEMENT OF THE TEST AND EVALUATION OF EVOLUTIONARY ACQUISITIONS

Co-chairs: Dr. Hank Dubin
Chief Scientist
US Army Operational Test and Evaluation Command

Mr. Don Greenlee
SAIC

Rapporteur: Dr. Stuart Starr
MITRE Corporation

Evolutionary acquisition is a strategy for the acquisition of systems which seeks to obtain and deploy system capabilities incrementally in time so that the user may enjoy the benefit of partial system capability as soon as possible without waiting for complete system implementation. Controversy over the merit of this approach within and among the affected communities exhibits a breadth of polarity ranging from assertions that no evolutionary acquisition has ever been successful to the insistence that all successful acquisitions have been inherently evolutionary. The focus of this working group was the (test and) evaluation of systems acquired under the evolutionary strategy, including the challenge of performing adequate evaluation of such systems prior to their deployment. This report describes working group **objectives**; the **approach and conduct** of its activities; major **issues, findings and observations**; and the group's **recommendations**.

OBJECTIVES

The following were taken as the principal objectives of the group's efforts:

- Develop a definition of "evolutionary acquisition."
- Develop a definition of "evaluation."
- Generate a list of characteristics of systems which are **suited** for evolutionary acquisition and a list of characteristics of systems **unsuited** for evolutionary acquisition.
- Identify evaluation approaches appropriate for systems acquired under the evolutionary strategy.
- Assess the compatibility of evolutionary acquisition with DoD acquisition regulations and public law.

- Articulate the external challenges to the successful application of evolutionary acquisition.

In addition to the above formal objectives, the group took for study and discussion the following related topics:

- Special considerations in the development and articulation of operational requirements peculiar to systems acquired under the evolutionary strategy.

- Measures of effectiveness, measures of merit, measures of performance, etc. appropriate for systems acquired under the evolutionary strategy.

- Obstacles and constraints to the successful application of evolutionary acquisition.

- Risks and benefits attendant on the use of the evolutionary strategy.

APPROACH AND CONDUCT

Members of the group participated in an initial introductory discussion of evolutionary acquisition and evaluation which touched on a variety of topics, including context, history and examples. The objectives and ancillary study topics were stated and explained, and the group divided into subgroups to address each one. After several sessions of full group and subgroup activity, findings and observations were presented by each subgroup on its assigned topic. These were then synthesized to generate the working group's product.

ISSUES, FINDINGS, AND OBSERVATIONS

The definition of "evolutionary acquisition" developed by the group in attainment of its **first objective** is as follows:

> **Evolutionary acquisition** is an acquisition strategy under which a general overall vision of mission need is satisfied by the fielding of an initial core capability, followed by incremental enhancements to meet evolving requirements based on feedback from the development, test, evaluation, user and technology communities.

In addition, it was noted that the application of evolutionary acquisition entails three fundamental activities:

- Formulation of a broad vision for the mission need at the outset considering general requirements and an architectural framework

- Development of an initial core capability having specific requirements with sufficient military utility for fielding

- Sequential upgrades of requirements, and the development and fielding of subsequent increments.

Furthermore, the group observed that the basic motivation and justification for evolutionary acquisition is a recognition that, for systems properly acquired under this strategy, the requirements will be subject to

frequent upgrades, refinements or clarifications due to one or more of the following:

- User/supporter feedback from the field
- Technology "push"
- Lessons learned from test and evaluation
- Changing missions or doctrine.

In meeting its **second objective**, the group developed the following definition for "evaluation":

> **Evaluation** is an independent assessment of mission effectiveness and suitability based on a wide range of information to support acquisition or deployment decision-making.

In order to accomplish its **third objective**, the group listed the following **characteristics typical of systems suited for evolutionary acquisition**:

- The detailed long-term requirements for the system are not well defined
- Complete system implementation requires technology solutions which are not yet in hand (but promise to be)
- Production units of the system are high-cost and few-of-a-kind
- The system's operation involves extensive human interfaces
- The system must adapt to changing command structure and tactics
- The system has complex interoperability and interdependency requirements with other systems
- Significant design changes are anticipated for the system
- There are high-technology and commercial-off-the-shelf insertion opportunities in the system development
- The system has inherent flexibility and growth capability ("evolvability").

Automated information systems, command and control systems and fusion systems are types of systems which are generally most amenable to evolutionary acquisition.

Next, the group listed **characteristics typical of systems unsuited for evolutionary acquisition:**

- The system has specific, well-defined requirements for its entire foreseeable lifetime
- Solutions to the system's design problems are accessible through existing technology
- Evolvability of the system cannot be demonstrated or proved
- The system's functionality is achievable through relatively simple "mechanical" means

- Large production runs of many system units are planned
- System functionality implemented by software is achievable via closed-form algorithms
- Significant basic design changes in the system are not anticipated.

Furthermore, pre-planned product improvement (P3I) and "block" improvements are acquisition strategies different from, and not to be confused with, evolutionary acquisition. Likewise, incremental acquisition of system units forced by spending constraints and funding limitations does not, by itself, constitute evolutionary acquisition, absent the hallmarks of incomplete requirements definition or uncertain technological solutions.

In addressing its **fourth objective** relative to the evaluation of systems acquired under the evolutionary strategy, the group observed that a number of general considerations should guide the choice of evaluation approaches. The overall evaluation should

- Use all appropriate data from:
  - Field/operational tests
  - Technical/engineering tests
  - Modeling and simulation
- To provide independent information on
  - Risk (cost, schedule, performance)
  - Operational suitability
  - Effectiveness (mission accomplishment/utility)
- To support the needs of
  - Acquisition decision-makers
  - System architects and developers
  - Users and supporters
- With respect to
  - The existing baseline
  - Requirements and specifications
  - Development/procurement/fielding
- Considering
  - Conditions other than those tested
  - Improvements in procedures and operational concepts
  - Technology growth and transfer.

With these considerations in mind, the group went on to **identify evaluation approaches appropriate for systems acquired under the evolutionary strategy**:

- Continuous Evaluation

- Periodic assessments
- Feedback from field use and training
- Threat updates

- Risk Assessment
  - Growth potential
  - Computer resource utilization
  - Software extensibility
  - Interoperability protocols and standards
  - Technology forecasting

- Well-planned Test Program
  - Baseline of system capabilities
  - Maturity matrix
  - Testbed
  - Drivers and instrumentation
  - Performance envelope exploration

- Continuity of Measures
  - Ability to meet mission need
  - Appropriateness to field use

- Feedback
  - User surveys
  - Modeling and simulation
  - Requirements engineering.

The group's **fifth objective** was to **assess the compatibility of evolutionary acquisition with DoD acquisition regulations and public law**. To this end, a survey was made of the recently issued DoD 5000-series of acquisition guidance documentation (DoD Directive 5000.1, DoD Instruction 5000.2 and DoD Manual 5000.2-M). These documents actually say very little about evolutionary acquisition. In fact, the term "evolutionary acquisition" is found in only two places in DoDI 5000.2:

> "Alternative acquisition strategies include **evolutionary acquisition** and preplanned product improvement."
> (DoDI 5000.2, 5.A.3.e)

> "**Evolutionary acquisition** is an approach in which a core capability is fielded, and the system design has a modular structure and provision for future upgrades and changes as requirements are refined. An **evolutionary acquisition** strategy is well suited to high technology and software intensive programs where requirements beyond a core capability can generally, but not specifically be defined. This approach is described in Joint Logistics Commanders Guidance, *Evolutionary Acquisition, An Alternative Strategy for Acquiring*

Command and Control Systems'." (DoDI 5000.2, 5.A.3.e.(1)

The referenced Joint Logistics Commanders Guidance contains its own definition:

> "**Evolutionary acquisition** is an acquisition strategy which may be used to procure a system expected to evolve during development within an approved architectural framework to achieve an overall system capability. An underlying factor in evolutionary acquisition is the need to field a well-defined core capability quickly in response to a validated requirement, while planning through an incremental upgrade program to eventually enhance the system to provide the overall system capability. These increments are treated as individual acquisitions, with their scope and content being the result of both continuous feedback from developing and independent testing agencies and the user (operating forces), supporting organizations and the desired application of new technology balanced against the constraints of time, requirements, and cost."

It is noted that the three definitions (JLC, DoDI 5000.2, and the group's) are not contradictory, but rather complementary. Another observation is that evolutionary acquisition is simply one of several alternative acquisition strategies, and that alone does not substantively alter the regulatory guidance for the acquisition process. The distinction between evolutionary acquisition and pre-planned product improvement is basically that only the near-term requirements in an evolutionary acquisition can be stated with specificity and that long-term requirements can be stated only generally, thus necessitating their refinement in the future.

The distinction between evolutionary acquisition and <u>all</u> other acquisition strategies is that under evolutionary acquisition, a conscious, public decision is made early in the program to develop a system incrementally, and to refine the requirements as increments of the system that can be developed, tested, evaluated and fielded, and, as technology matures, to provide solutions to these requirements, if necessary. This eliminates the need for a Milestone IV decision point to assess the requirement for restarting an acquisition program for a system upgrade. Each "iteration" in an evolutionary acquisition is to be treated as a separate acquisition beginning with approval of the refined requirements at Milestone 0, although waivers of the Phase 0 program reentry may be granted by the milestone decision authority.

In summary, the DoD 5000-series guidance pertaining to test and evaluation of systems acquired under an evolutionary strategy can be stated quite simply: such systems are to be treated no differently from those acquired under any other strategy, except that they are expected to go through more than one acquisition cycle. As a

consequence, it will be necessary to repeat the acquisition phases and successfully pass each of the acquisition milestones as though successive developmental iterations were, in fact, totally different acquisition programs. The actual number of acquisition phases and milestones to be repeated is to be determined by the milestone decision authority.

The group's **sixth and final objective** was to **articulate the external challenges to the successful application of evolutionary acquisition**. The potential motivations and justifications for employing the evolutionary acquisition strategy have been discussed above, as has the fact that no DoD acquisition regulations prevent its use. The principal general external challenges to its successful application appear to be the following:

- Uncertainty in the development community as to its utility
- Mistrust in the oversight community of its "legality"
- Discomfort in the test and evaluation community regarding how to carry out their responsibilities.

Discussions with program managers in government and industry suggest that some are unpersuaded as to the benefits of evolutionary acquisition, if they are aware of it as a potential tool at all. Developers who viewed evolutionary acquisition favorably felt it provided them with flexibility in achieving their capability objectives, whereas the test and evaluation community generally believed that this strategy promoted a lack of discipline.

The second point has to do with the circumstance that some with program review responsibilities regard evolutionary acquisition with suspicion, regarding it as a means by which developers might seek to avoid the articulation of system requirements and the conduct of adequate test and evaluation. The greatest challenge to evaluators is in assessing the risk arising from uncertainty as to whether the early design or architecture with limited functionality can eventually support the full functionality as established in the system's capability objectives.

Lastly, there has been little discussion by testers and evaluators as to methods for T&E in an evolutionary environment. An obvious problem here, for example, is in regression testing, i.e., determining the amount of retesting to be performed as new increments of capability are added to the core system.

In regard to institutional challenges, the existing general definition(s) of evolutionary acquisition do not provide crisp criteria for deciding whether a particular system is appropriate for evolutionary acquisition; existing DoD and Service regulations do not provide adequate guidance in this area and, under some interpretations, inhibit or even preclude evolutionary acquisition as a strategy. Given the eventual documentation of criteria for the use of evolutionary acquisition and regulations for its application, it will still be necessary then to educate all elements of the acquisition community on the theory and practice of implementation. This process could be aided by a comprehensive, critical review of past and ongoing programs considered to have employed the evolutionary acquisition

strategy, both successfully and unsuccessfully. In view of the iterative loop nature of the evolutionary acquisition process, another challenge will be to establish the crucial existing system baseline and the core capability baseline by which program progress can be measured.

With respect to implementation challenges, a basic one arises directly from the inherent flexibility which makes evolutionary acquisition an attractive strategy in appropriate cases, namely, controlling this flexibility by establishing a sufficiently disciplined approach so that the program is prevented from drifting off the path of progress. Another challenge is the need for the early identification of the essential, militarily useful core capability which will serve as the nucleus for further evolution. The evolutionary acquisition concept envisions the possibility of multiple parallel developments where, for example, full scale development of the core capability may be taking place concurrently with conceptual or demonstration work on the next increment of system capability. Providing the resources for these concurrent activities constitutes another challenge. A system design challenge arises from the need to build in sufficient flexibility such that evolution is promoted rather than hindered and no potentially desirable evolutionary paths are prematurely blocked.

Feedback is another critical element in the evolutionary acquisition concept; a disciplined, visible mechanism has to be established and exercised to ensure that information from testing and operational experience flows upstream and properly impacts system evolution. In addition to affecting the design of the current system increment, this feedback should also play a major role in the requirements refinement process.

To assist both in formal test and evaluation and in informal collection of experimental data, the creation of an exercise and test environment is usually called for. This environment typically includes a prototype of the evolving system, simulators, stimulators, replicas of interoperating (or adversarial) systems, instrumentation and users or their surrogates. Adequate funding to create and maintain this environment is yet another challenge, especially since there may be a need to procure multiple systems in order to create an operationally realistic critical mass. A significant challenge for evolutionary acquisition is the early definition and faithful implementation of an adequate and affordable program of experimentation, test and evaluation. Experience has shown that program managers often have difficulty in preventing the projection of a "test avoidance" image as the program proceeds along its evolutionary path. A generalization of this challenge is equally serious, namely, that of keeping the overall program goals in view and escaping diversionary loops and tangents harmful to the overall evolutionary process.

A final challenge noted is that of selection of the proper contract type. Some forms of contract vehicle are obviously inappropriate for an evolutionary acquisition, e.g., a fixed price contract for the entire system (although fixed price may be acceptable for an individual system increment). Another consideration is the maintenance of competition among contractors, since an incumbent vendor may have a significant advantage in bidding for the next increment(s).

The group's discussion of its study topics led to some observations on requirements definition, risks and benefits, and test and evaluation. Particular considerations for systems to be acquired under an evolutionary strategy include:

- the mission area architecture—the need to establish goals for ultimate operational capability;
- the identification of specific core requirements (the first phase must be clearly defined as a baseline for providing a new and effective capability to the user);
- interoperability requirements among potential interfacing systems and among using organizations, e.g., the Services;
- modular and evolvable design—a building block approach in which each new increment mates smoothly with the previous one; and
- planned "obsolescence" or awareness of relevant technology developments and anticipation of additional capability as the system matures.

The actual generation and documentation of requirements should involve the balanced contributions of all the traditional communities. These include users to articulate operational needs; designers, technologists and system engineers to provide a vision of what is technically achievable; test and evaluation experts to address testability—both developmental and operational; and program and policy personnel for guidance and resource inputs. This team must be able to translate operational requirements into technical specifications, carefully maximizing the application of useful technology without distorting the original statement of need. Measures of performance and effectiveness must also be defined to grade system progress toward full effectiveness and suitability, while distinguishing between threshold and mature capabilities.

In discussing the potential benefits associated with evolutionary acquisition, the group noted the opportunity to field system capabilities early in the face of requirements and technology uncertainties, the advantage of user feedback in developing future requirements, and the possibility of overall cost and time savings. However, the attendant risks include:

- the possibility of developing requirements which are overly (or insufficiently) ambitious or insufficiently (or overly) specific;
- defining individual increments which do not support or converge to the ultimate goal (loss of the "big picture");
- inadequate user/developer communications to permit the evolution of requirements;
- failure in one increment jeopardizing the long-term objective;
- evading, or projecting the appearance of evading, responsibilities for conducting adequate test and evaluation due to the intrinsic "moving target" nature of evolutionary acquisition; and

- closer scrutiny and greater reluctance to approve by the oversight community because of the novelty of the approach.

The inherent structure of the evolutionary strategy, though, is comparatively conservative, in that risk is faced in small increments which can be assessed and dealt with, rather than in a total program, all-or-nothing fashion.

Regarding test and evaluation considerations, the group observed that the basic concept in evolutionary acquisition is to "build a little, test a little, learn a lot". Tools with which to bridge the gap between the user and the developer are especially important, because of the critical need for feedback on system performance and evolving requirements. A plan for developing test capability needs to be prepared early in the program in order to obtain resources and to ensure the test capability will be in place when needed. The plan should address flexibility in design and be adaptive to the system's growth; have rapid prototyping features to allow demonstration of new operational capabilities; possess growth ability parallelling that of the system itself; and the difficult verification and validation function (in general, better methods are needed). A structured, rather than ad hoc, interface among users, testers and logisticians should be defined.

Finally, an evaluation plan should be prepared which permits the extraction of operational functions from the requirements. For each function, measures of effectiveness and suitability must be defined; this may be a difficult task, but a necessary one to provide a consistent baseline for evaluation. These measures are to be evaluated using all available test data, including field, laboratory and simulation (but appropriately separated). The plan should also possess the flexibility to add new measures to accommodate new requirements as they are defined, while ensuring that they are consistent with existing ones.

RECOMMENDATIONS

Evolutionary acquisition is a strategy which has the potential for permitting the responsible discharge of duties toward articulating system requirements and conducting adequate test and evaluation without postponing the joy of early deployment and incremental procurement of useful capabilities. For these possibilities to be realized, however, great care must be taken in structuring the system development program, beginning with a demonstrable determination that the system is capable of evolution, i.e., that technological solutions for achieving the early increments of capability are not dead ends which will block further growth to the ultimate desired system objectives. With this assurance in hand, evolutionary acquisition actually becomes a risk reduction strategy, in that commitment is made sequentially to relatively small increments of achievement, rather than gambling on the ability to accomplish a single ultimate goal with no interim useful phases. And, of course, the acquisition review community must be persuaded of the prudence of embarking on an enterprise whose long-term requirements and solutions are only generally defined.

Given that an evolutionary approach is chosen, several elements are recommended to minimize risks, maximize benefits and

allow test and evaluation to play its proper role in disciplining the program. As noted in the findings under objective number four above, these include a continuous evaluation process, an ongoing risk assessment methodology, a well-planned test program, continuity in measures of performance and effectiveness and a robust feedback mechanism. In particular, the overall evaluation should use all appropriate data to provide independent information regarding risk, effectiveness and suitability to acquisition and deployment decision-makers, developers and users. This data, based on field/operational tests, technical/engineering tests and the results of modeling and simulation, should relate to the existing baseline, requirements and specifications, and development/procurement/fielding, and should establish a foundation for the consideration of conditions other than those tested, improvements in procedures and operational concepts, and technology growth and transfer.

WORKING GROUP V PARTICIPANTS:

NAME	ORGANIZATION
Maj Jerald A. Coleman	HQ AFOTEC/LG5
Dr. Philip C. Dickinson	E-Systems, Inc.
Dr. Henry C. Dubin	US Army OPTEC
Mr. Raymond G. Fontaine	US Army TECOM
Mr. Robert C. Good	General Dynamics
Mr. Donald R. Greenlee	SAIC
LTC Jimmy R. Grussmeyer	ODUSA(OR)
Ms. Jacqueline Hamilton-Amthor	ITEA
Mr. Richard E. Helmuth	Douglas Aircraft Company
Mr. Ervin Kapos	Kapos Associates
Dr. Wayne R. Knight	US Army OPTEC
Col. Thomas Y. Kumashiro	HQ AFOTEC/TES
Mr. Thomas E. Nolan	US AMSAA
Mr. Patrick J. O'Neill	US AMSAA
Dr. Nelson S. Pacheco	Institute for Defense Analyses
Mr. David W. Prichard	US Army TECOM
Dr. Stuart H. Starr	The MITRE Corporation
Dr. Alfred E. Victor	Institute for Defense Analyses

WORKING GROUP VI

USING TESTING TO ENHANCE THE CREDIBILITY OF OUR ANALYSIS TOOLS

Co-chairs: Mr. Allen Murashige
Air Force Studies and Analyses Agency

Mr. Dave Hall
Naval Weapons Center

Rapporteur: Mr. Vern Bettencourt
MITRE Corporation

MISSION STATEMENT AND OBJECTIVES

The mission of Group VI was to explore ways in which the tester can assist the evaluator, in particular, methods of using test results to improve analytical tools (ie, computer models) in order to raise our level of confidence in their credibility. The objectives of Group VI were developed from the mission statement:

- To explore methods of using test results to improve analytical tools;
- To identify critical issues to be addressed in trying to use testing to improve analytical tools;
- To obtain different perspectives on these issues through discussion, sharing experiences and insights;
- To surface areas of agreement/disagreement with regard to what is necessary to achieve the stated mission;
- To derive recommendations which would facilitate the achievement of the mission.

BACKGROUND

In recent years Congress and the Office of Secretary of Defense have become increasingly aware of the pivotal role that computer models play in analysis, evaluation, and decision making during the development of modern weapon systems. In particular, concern was raised on the credibility of such models in the 1988 GAO study, "DoD Simulations: Improved Assessment Procedures Would Increase the Credibility of Results," the 1989 Defense Science Board study, "Improving Test and Evaluation Effectiveness," and the 1989 OSD/DOT&E document, "Policy for the Application of Modeling and Simulation in Support of OT&E."

Congress' and OSD's concerns have been manifest in new regulations which require computer simulations used in major weapon system decisions to be validated and

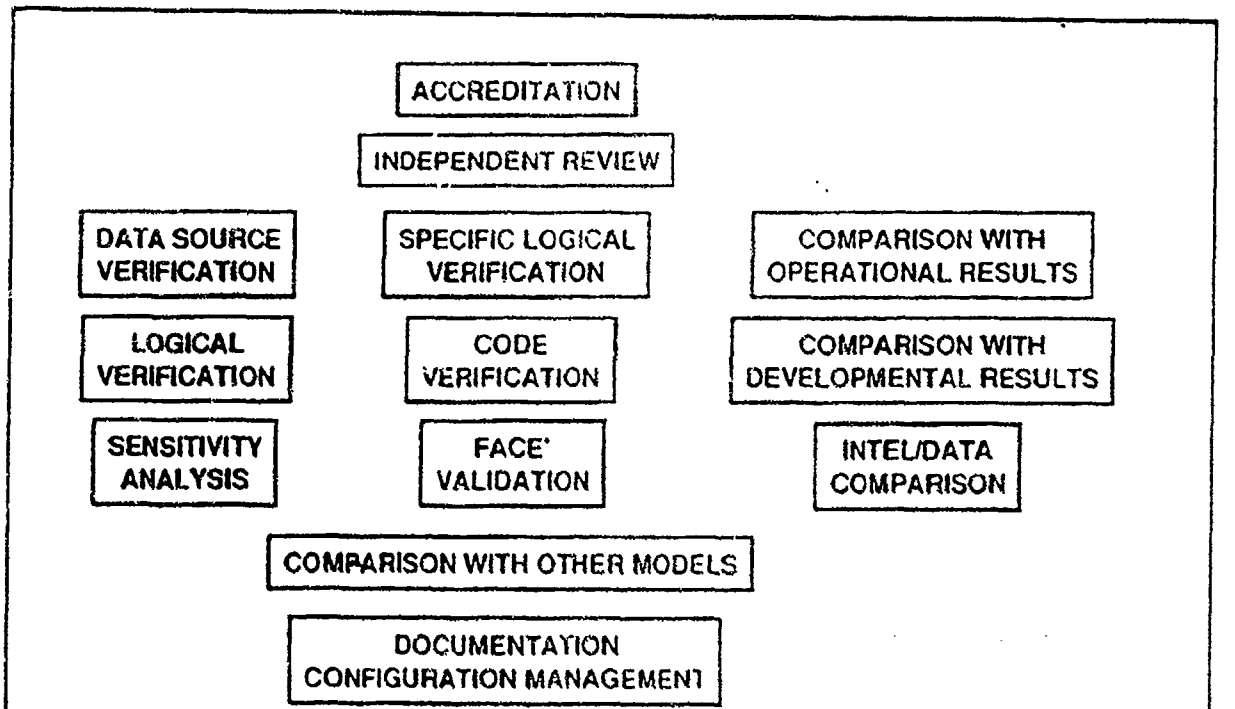


FIGURE 1. CONTRIBUTORS TO MODEL CREDIBILITY

accredited, and in the creation of the Defense Modeling and Simulation Office, which received generous funding from Congress and is currently being organized and staffed by OSD and the Services. Other examples of the added attention being paid to model credibility are the Joint Technical Coordinating Group for Aircraft Survivability (JTCG/AS) model validation program called "SMART," and the DoD CROSSBOW-S Joint Modeling and Simulation System threat model development and validation.

This emphasis on enhancing the credibility of the models used in weapon systems analyses will necessitate improved, institutionalized processes to establish model credibility. A fundamental aspect of this credibility enhancement process is model validation, and one of the most convincing methods of validation is through comparison of model predictions with open air hardware test results. While difficult to accomplish, this model T&E process can provide essential feedback on model performance which identifies model aspects needing refinement and acts as an external check on the ability of the model to simulate real world conditions.

As a result of the attention being paid to model credibility, MORS has sponsored the SIMVAL symposia, working to achieve

a consensus within the analytical community on the terminology and procedures surrounding the establishment of model credibility. The current MORS definitions of the primary terms are:

- **Verification** is the process of determining that a computer model implementation accurately represents the developer's conceptual description and specifications.

- **Validation** is the process of determining the degree to which a model is an accurate representation of the real-world from the perspective of the intended uses of the model.

- **Accreditation** is the official determination that a computer model is acceptable for a specific purpose.

With regard to the process of establishing model credibility, there are several frequently cited contributing elements shown below in Figure 1. MORS is sponsoring a workshop in April 1992 to develop a monograph which will detail the procedures to accomplish each part of this credibility building process.

The elements in this process contribute to a model's credibility. All items are not essential for model accreditation; however, the more complete the set of components, the stronger will be the model's credibility. Comparison with operational and developmental test results (which is the focus of this discussion group) provides perhaps the most convincing evidence to build credibility, especially if there is good correlation between the model's predictions and corresponding real (open air) events which are observed within a controlled/ instrumented environment. Therefore, under the appropriate conditions, data from weapon system tests can be very valuable for enhancing model credibility, and the test community and the weapons program offices (both of whom rely extensively on models needing validation) should be made to understand and support the use of test data for this ancillary purpose.

THE ISSUES

After preliminary deliberations, the group agreed to structure their discussion around the following list of questions addressing the use of testing to enhance model credibility.

Why...

- Why should we be concerned with model credibility?

- Why should we be trying to use test results to enhance model credibility?

- Why must we do anything at all?

What...

- What actions are required to facilitate the use of tests?

- What must be considered if testing (or use of test data) is proposed to validate a particular model or set of models?

- Is (open air) testing necessary or useful?

- What are the test constraints and artificialities?
- What must be achieved? What are the test objectives?

Who...

- Who should be responsible?
- Who are the key players?

When...

- When must you initiate/complete the process?

How...

- How do you accomplish what it is that you need to do?
- How do you obtain the needed resources (money, manpower, data, etc)?
- How much should you attempt and how well can you achieve it?

DISCUSSION

The members of Group VI represented a broad cross section of the DOD and defense industry, including members from the T & E community, the operational community, the analysis/modeling community, and the weapon system acquisition/development community (see attached participants list). Members of the group had differing degrees of exposure/experience in the area of model validation through testing. Group members presented short briefings on a sampling of programs completed, underway or planned within DOD which address enhancing model credibility:

- US Army Model-Test-Model Methodology
- US Air Force Model Accreditation Plan for the F-22 (ATF)
- JTCG/AS Susceptibility Model Assessment with Range Tests (SMART)
- US Army AH-64 OT-II Crew Performance Analysis
- US Air Force Advanced Medium Range Air-to-Air Missile (AMRAAM) Testing and Analysis

It was observed that the Services have had varying degrees of success in developing, applying and accrediting computer models, with some branches providing a better institutional environment (regulation, control, and oversight) for this process than others. As a whole, however, the DoD has been inconsistent as to the level of discipline governing the application of computer models in weapon system development.

The "Why's"

The "Why" questions were the least contentious and time consuming to address. The opening presentation and ensuing discussion provided the background and established the need and interest in using testing/test data to enhance model credibility. The publication of the GAO, DSB, and OSD/DOT&E reports were cited as events which raised the level of concern for model

validity/credibility in Congress and OSD; the reader is urged to review these documents if he is not already familiar with them. The use of models within the DoD will face more stringent control as a result of new regulations requiring the validation and accreditation of models affecting major weapon system acquisition decisions, as well as the establishment of the Defense Modeling and Simulation Office which will provide oversight and standardization in the use of models within the DoD.

Although it is widely accepted that model credibility needs improvement, and that testing may contribute to this process, it is also recognized that the toughest challenge to its success is getting the many separate camps who hold the key resources to cooperate. The camps are special interest groups defined and constrained by Service barriers, program barriers, classification barriers, and functional barriers (testing vs training vs planning vs development vs analysis). To successfully marry testing with model validation, these groups must recognize the value and understand the objectives of the program. They must be willing to sacrifice a bit of autonomy, perhaps readjust their schedules, loosen their hold on test data, and work together on the program.

The "What's"

The breadth of experience represented in the group led to numerous valuable insights on what and what not to do to promote the use of testing in model validation. Perhaps the most commonly held view was that two essential ingredients are: 1) informing senior management about modeling and simulation VV&A needs, and obtaining strong, early commitment of support, and 2) early and continual coordination among the key players from the analysis/modeling, testing, and weapon system program offices. In order to accomplish the latter, modeling and simulation applications (which models are being used, what are they being used for, what are the most critical modeling parameters) and model validation/testing requirements should be identified by analysts and test evaluators as soon as possible. These model testing needs (including resources) should be identified and emphasized within the TEMP. Modeling and simulation working groups should be formed to coordinate the VV&A needs with test planning. These working groups should involve members working with the COEA, to incorporate their VV&A needs and feed back model validation information into the COEA process.

The key to the eventual success of M&S credibility enhancement will be the publication of VV&A results for peer review and archiving. The neglect of this practice, a discipline which should be fundamental to establishing any scientific credibility, is a leading cause for skepticism among the model critics. Where test data is used in model validation, this necessitates release of data at some level. Moreover, the process should be institutionalized so that the VV&A process is not resident within one program but is perpetuated over the life of the model, across its span of applications. The new DOD Modeling and Simulation Office could play a vital role in institutionalizing the VV&A process within the DOD.

The "Who's"

Responsibility for coordinating modeling and simulation VV&A practices

within the DoD should be assumed by the DoD Modeling and Simulation Office. MORS is the logical agency for developing the professional (analysts, modelers, testers, etc) community's recommendations on what these practices and procedures should be. Each of the Services should select an office to be responsible for establishing VV&A policies and procedures (including the use of test data for model validation). Possible offices might be the Deputy Under Secretary (Operations Research) for the Army, COMOPTEVFOR for the Navy, and AFOTEC for the Air Force. In order to encourage the use and sharing of test data for model validation, it appears that the agencies responsible for testing (such as AFOTEC or COMOPTEVFOR) may be able to influence the key players in the community (including the weapon system program offices) to support and participate in this effort. Multi-service offices such as the Joint Technical Coordinating Group or the DMSO should promote, sponsor, archive and distribute model VV&A technical reports.

The "How's"

To increase momentum behind the model VV&A process, it will be necessary to educate members of the acquisition community (program managers, decision makers, analysts, evaluators, testers, modelers) of the importance, benefits, and costs of the M&S VV&A process, noting the increased attention from Congress and OSD on model credibility. Courses on the applications of analysis and models in weapon system development should be introduced at service schools and the DSMC. It will also be necessary to educate the analysts and modelers about how to effectively verify and validate their models with limited resources and test data, and how to effectively integrate their validation needs into a weapon system test plan with minimum disruption. MORS should continue to develop model VV&A concepts and procedures and explore the design of model validation experiments (tests). MORS or perhaps SURVIAC could sponsor short courses or symposia on these topics. The Services should develop standard VV&A procedures and establish an office of primary responsibility for model VV&A policy. The DMSO should promote and fund efforts to coordinate VV&A procedures among the Services.

The "When"

Analysts, modelers, and program managers should start considering model VV&A needs as soon as models are identified for use in the program (whether for requirements analysis, development, or testing). Testing to support model validation needs should be planned and integrated as the TEMP is being developed. Model testing and validation will be complete for a program when the accreditation decision has been made (accreditation should occur prior to major decisions influenced by the model), however the validation and testing process should continue for a model as long as it is in use.

FINDINGS

1. Within the DOD, there have been several efforts to use testing and test data to validate models in order to enhance their credibility, which have achieved varying degrees of success. Obtaining useful/meaningful test results is difficult, and validation is usually constrained to only a narrow domain of conditions. Emphasis and effort directed to

the validation of models (especially through testing) is inconsistent among the services.

2. There is a lack of well defined processes and standards in DOD governing model verification, validation, and accreditation, as well as the use of testing to support validation. Requirements to document, publish, and archive model validation test results would help build credibility.

3. With the development of complex 'smart' weapon systems with embedded sensors and integral C3, emphasis is being placed on evaluating weapons in more complex scenarios. At the same time, testing budgets and resources (e.g., airspace) are becoming tighter. These factors are leading to more reliance on digital simulations and in particular, higher level digital simulations (many vs many play instead of one-on-one) in the test and evaluation process. Besides being more difficult to develop, these higher level simulations are more difficult to validate.

4. There is more reliance on models and more emphasis on model credibility in requirements analysis, such as cost and operational effectiveness analyses and mission area analyses. This will demand more rigorous model VV&A practices.

5. Model VV&A using testing is easier when the weapon system programs are users of the model and concerned with its validity; it is very difficult otherwise. Data protection and possessiveness by program offices is a major hurdle to overcome, limiting the use of test data or model validation results outside the program.

6. Model VV&A should continue over the life of the model, to continually improve credibility and to validate model updates and modifications. Since models are frequently used by several weapon system programs as well as other users (eg, training, war planning), the release of data for publication of validation test results would provide efficiencies and help build model credibility for future use. In this era of vanishing resources, such efficiencies and cooperation will be mandatory if programs are to survive.

7. When analyzing test results and comparing test data to model output, there is a fine line between model validation and model calibration; this distinction should be further explored within the SIMVAL arena.

RECOMMENDATIONS

1. The verification, validation, and accreditation of models and simulations used by weapon system programs must be given greater attention within the DoD if the concerns of Congress and OSD are to be satisfied. The leadership of the weapons development community should be made aware of the resources--time, money, manpower, and data--that this process demands.

2. Where possible, model validation through comparison with results from (preferably open air) weapon testing should be attempted and supported. Model VV&A requirements should be identified and emphasized in the test and evaluation master plans. The program managers and testing community should be educated as to the benefits of developing validated, credible models. DOD (DMSO) should promote and support model VV&A efforts, especially those employing test data.

3. The modeling and simulation VV&A process should be coordinated and institutionalized in the Services. Again, the DoD DMSO could play an important coordinating role.

4. Technical reports describing M&S VV&A should be published and archived, preferably at multi-service organizations (eg, DMSO, JTCG). The DoD should work to reduce program and Service barriers to make testing and data more available for model VV&A use.

5. Modelers and analysts must be educated on effective VV&A techniques and the limitations of testing. Experimental design concepts must be employed to enable the validation of critical model capabilities with limited testing and data. Careful attention must be paid to testing conditions, sample sizes, and artificialities when applying test data to model VV&A.

WORKING GROUP VI PARTICIPANTS:

NAME	ORGANIZATION
Mr. Robert D. Aaron	US Army OPTEC
Mr. Vernon M. Bettencourt	The MITRE Corporation
LT Bradley S. Brockmeyer	OPTEVFOR
Mr. Jesus E. Carrillo	Research Analysis and Maintenance, Inc.
Mr. Wayne E. Deutscher	COLSA, Inc.
Ms. Sylvia A. Diaz	HQDA, OASA(RDA)
MG James E. Drummond (ret)	Consultant
Dr. Eugene F. Dutoit	US Army Infantry School
Lt. Col. Michael J. Gough	HQ AFOTEC/ST
Mr. Donald A. Gray	US Army TECOM
Mr. David H. Hall	Naval Weapons Center
Mr. Eric L. Keck	HQ AFOTEC/OAN
MAJ James C. Kolding	US Army TRAC-WSMR
Mr. Warren M. Lee	HQ SAC/XPSS
Mr. Robert H. MacFadzean	Naval Surface Warfare Center
Mr. Allen M. Murashige	HQ USAF/SAG
Mr. Chris Pfledderer	ASD/ENSA
Dr. Adelia E. Ritchie	SAIC
Mr. Tom Severson	Naval Weapons Center
Lt. Col. Bruce L. Smith	AFCSA/SAZ
Ms. Lounell D. Southard	US Army TRAC-WSMR
Mr. Clayton J. Thomas	HQ USAF/AFSAA
LTC Jesse R. Tucker	Defense Information Systems Agency
Dr. Stewart W. Turner	WNG Associates, Inc.
Dr. Clairice T. Veit	The RAND Corporation
Dr. Steve Warner	Institute for Defense Analyses
Mr. Ronald L. Wood	Research Analysis and Maintenance, Inc.

ITEA/MORS MINI-SYMPOSIUM

EMPHASIZING THE "E" IN T&E

AGENDA

MONDAY, 30 SEPTEMBER 1991

1830-2000: Early registration and Car Pass Pick-Up

2000-2100: Chairmen/Rapporteur's Meeting

TUESDAY, 1 OCTOBER 1991

0700-0800: Registration

0800-0810: Opening Remarks
General Chairman, **Mr. Edward Brady**

0810-0820: Societies Welcome
Mr. Vern Bettencourt, MORS President
Mr. Matt Reynolds, ITEA President

0820-0830: Host Welcome
CAPT S. Buescher, Director, Wargaming Center
Naval War College

0830-0930: Keynote
Mr. Walt Hollis
Deputy Under Secretary of the Army (Operations Research)

0930-1000: Break

1000-1200: Decisionmakers' Challenge to Working Groups (Panel)
Technical Chairman, **Dr. Patricia Sanders**

1200-1330: Lunch

1330-1630: Initial Working Group Sessions

1730-2100: Clambake

WEDNESDAY, 2 OCTOBER 1991

0800-1130: Working Group Sessions Continue
Coffee Break (flexible)

1130-1300: Lunch Break (flexible)

1300-1700: Working Group Sessions Continue

1700-1900: Chairmen/Rapporteur's Meeting

THURSDAY, 3 OCTOBER 1991

0800-1000: Final Working Group Session

1000-1030: Break and Chairmen's Meeting

1030-1200: Plenary Wrap-up Session (Technical Chairman)
Summaries
Recommendations

1200-1300: Chairmen/Rapporteur's Meeting

1300-1500: Tour of Naval Underwater Systems Center (NUSC) (optional)

PARTICIPANTS

Aaron, Robert D.
US Army Civilian

US Army Operational Test & Evaluation Command
ATTN: CSTE-ECC-Z
4501 Ford Avenue, #920
Alexandria, VA 22302-1458
(703) 756-2306

Addison, Natalie S.
Commercial

Military Operations Research Society
101 S Whiting Street
Suite 202
Alexandria, VA 22304-3483
(703) 751-7290

Apicella, Frank J.
US Army Civilian

US Army Operational Test & Evaluation Command
Park Center IV, ATTN: CSTE-CS
Alexandria, VA 22302-1458
(703) 756-0380

Auletta, Joseph F.
LtCol, USAF

AFSC/INA
887 Grove Hill Road
Wright-Patterson AFB, OH 45433
(513) 257-4751

Ball, Robert E.
US Navy Civilian

Naval Post Graduate School
Code AA/Bp
Monterey, CA 93943
(408) 646-2885

Bates, Michael D.
USAF Civilian

OAS(AFSC)
3608 St Andrews Drive
Kirtland AFB, NM 87117-6008
(505) 844-0231

Bean, Theodore T.
FFRDC

The MITRE Corporation
Washington C3 Center
7525 Colshire Drive
McLean, VA 22102
(703) 883-6231

Bellach, Tony R.
USAF Civilian

USAF/Office of Aerospace Studies (AFSC)
Kirtland AFB, NM 87117-6008
(505) 846-8302

Berg, David H.
Maj, USAF

HQ, SAC/XPSW
901 SAC Blvd, Suite 2B5
Offutt AFB, NE 68113-5330
(402) 294-4220

Bettencourt, Vernon M. FFRDC

The MITRE Corporation
W 451
7525 Colshire Drive
McLean, VA 22102-3481
(703) 883-7364

Blackman, Wildon C. US Navy Civilian	Naval Weapons Center Warfare Analysis Program Code 304 China Lake, CA 93555 (619) 939-3833
Bolin, Stanley F. US Army Civilian	US Army Operational Test & Evaluation Command ATTN: CSTE-ESE-M 4501 Ford Avenue Alexandria, VA 22333 (703) 756-2166
Bors, Linda J. USAF Civilian	HQ SAC/XPSM Suite 2B5 901 SAC Blvd Offutt AFB, NE 68113-5330 (402) 294-7034
Boyles, Irvin E. US Army Civilian	US Army Operational Test & Evaluation Command 4501 Ford Avenue Alexandria, VA 22302-1458 (703) 756-2166
Brady, Edward C. Commercial	Strategic Perspectives, Inc. 7704 Lakeloft Court Fairfax Station, VA 22039 (703) 250-6338
Bridgewater, Gary Commercial	SAIC 8619 Westwood Center Drive Suite 608 Vienna, VA 22182 (703) 749-5658
Bridgman, Michael S. FFRDC	Logistics Management Institute 6400 Goldsboro Road Bethesda, MD 20817 (301) 320-7248
Brockmeyer, Bradley S. LT, US Navy	DEPCOMOPTEVFORPAC NAS, North Island San Diego, CA 92135-5115 (619) 545-2725
Brooks, Wilbert J. US Army Civilian	US Army Materiel Systems Analysis Activity ATTN: AMXSY-GA Aberdeen Proving Ground, MD 21005-5071 (301) 278-4608
Buescher, Stephen M. CAPT, US Navy	Naval War College Wargaming Dept Newport, RI 02841-5010 ( )
Carrillo, Jesus E. Commercial	Research Analysis and Maintenance, Inc. 1790 Lee Trevino Drive, #600 El Paso, TX 79936-4525 (915) 592-7047

Chernowitz, George Commercial	American Power Jet Co. 705 Grand Avenue Ridgefield, NJ 07657-1583 (201) 945-8203
Christie, Thomas P. FFRDC	Institute for Defense Analyses 1801 N. Beauregard Street Alexandria, VA 22311 (703) 845-2482
Claxton, John D. Other DoD	Defense Systems Management College Technical Management Department Fort Belvoir, VA 22060 (703) 664-5060
Coleman, Jerald A. Maj, USAF	HQ AFOTEC/LG5 Kirtland AFB, NM 87117-7001 (505) 846-5310
Collie, Allan L. LtCol, USAF	HQ AFOTEC/XPX 9607 Admiral Dewey, NE Kirtland AFB, NM 87117-7001 (505) 846-5242
Collier, Darrell W. US Army Civilian	US Army TRADOC Analysis Center-WSMR ATTN: ATRC-W White Sands Missile Range, NM 88002-5502 (505) 678-4512
Comfort, Gary C. FFRDC	Institute for Defense Analyses 1801 N. Beauregard Street Alexandria, VA 22311-1772 (703) 845-2245
Deutscher, Wayne E. Commercial	CCLSA, Inc. P.O. Box 1418 Belmar, NJ 07719-1418 (908) 280-1657
Diaz, Sylvia A. US Army Civilian	HQDA, OASA(RDA) ATTN: SARD-DOS Pentagon, Room 3E422 Alexandria, VA 22310-103 (703) 693-5441
Dickinson, Philip C. Commercial	E-Systems, Inc. (CAPA) 10530 Rosehaven Street, Suite 200 Fairfax, VA 22030-2886 (703) 352-0300
Douglas, Donald M. Col, USAF	HQ AFOTEC/OA Kirtland AFB, NM 87117 (505) 846-4209
Drummond, James E. Commercial	Consultant 2 Riding Path Hampton, VA 23669 (804) 851-5470

Dubin, Henry C. US Army Civilian	US Army Operational Test & Evaluation Command Park Center IV 4501 Ford Avenue Alexandria, VA 22302-1458 (703) 756-2367
Duff, James b. US Navy Civilian	COMOPTEVFOR Technical Director Code OOT Norfolk, VA 23511-5225 (804) 444-5197
Dufresne, Robert E. US Navy Civilian	Naval Underwater Systems Center Code 2092, Bldg 1259 Newport, RI 02841-5047 (401) 841-2722
Dutoit, Eugene F. US Army Civilian	US Army Infantry School ATTN: ATSH-CDC-O Fort Benning, GA 31905-5400 (404) 545-3165
Edwards, William C. Other DoD	National Security Agency ATTN: R562 9800 Savage Road Fort George G. Meade, MD 20755 (301) 688-0592
Ellner, Paul M. US Army Civilian	US Army Materiel Systems Analysis Activity ATTN: AMXSY-RM Aberdeen Proving Ground, MD 21005-5071 (301) 278-2090
Erickson, Terrence J. Commercial	Alliant Tech Systems, Inc. 600 Second Street, NE Hopkins, MN 55343 (612) 931-7817
Feighery, Daniel J. Commercial	Institute for Defense Analyses 1801 N. Beauregard Street Alexandria, VA 22311-1772 (703) 845-6974
[illegible] Bernard B. [illegible] Army	OSD DDR&E T&E Pentagon, Room 3E1060 Washington, DC 20301-3110 (703) 695-4608
Filanowicz, Robert W. Cdr, US Navy	Naval War College Wargaming Dept. Newport, RI 02841-5010 ( )
Finberg, Harvey J. Other Govt	US GAO NSIAD, Rm 5112 441 G Street, NW Washington, DC 20548 (202) 275-3935

Fontaine, Raymond G. US Army Civilian	US Army Test & Evaluation Command ATTN: AMSTE-TA-G Aberdeen Proving Ground, MD 21005-5055 (301) 278-5222
Fossett, Christine A. Other Govt	US GAO Office of Policy, Room 6800 441 G Street, NW Washington, DC 20548 (202) 275-1970
Foulkes, John B. US Army Civilian	HQ Department of the Army ATTN: DACS-TE Pentagon, Room 3C571 Washington, DC 20310-0102 (703) 695-8995
Frank, Richard T. LtCol, USAF	HQ AFOTEC/XPP 8811 Armistice, NE Kirtland AFB, NM 87117-7001 (505) 846-5237
Fries, Arthur FFRDC	Institute for Defense Analyses 1801 N. Beauregard Street Alexandria, VA 22311 (703) 845-2364
Garvey, Richard E., Jr. Commercial	BBN Systems & Technologies 33 Moulton Street Cambridge, MA 02138 (617) 873-4902
Gill, David J. Commercial	Rockwell International Tactical Systems Division 1800 Satellite Blvd Duluth, GA 30136 (404) 497-5196
Good, Robert C. Commercial	General Dynamics 12730 Soule Street Poway, CA 92064 (619) 542-4780
Gough, Michael J. LtCol USAF	HQ AFOTEC/ST 4820 B Cutting, NW Kirtland AFB, NM 87117-7001 (505) 846-0631
Grape, Eric US Army Civilian	US Army Materiel Systems Analysis Activity ATTN: AMXSY-LX Aberdeen Proving Ground, MD 21005-5071 (301) 278-6735
Gray, Donald A. US Army Civilian	US Army Test and Experimentation Command Methodology and Analysis Directorate ATTN: CSTE-TMA Fort Hood, TX 76544-5065 (817) 288-9966

Green, Frederick L. Col, USAF	HQ AFOTEC/TEZ Kirtland AFB, NM 87117 (505) 846-5130
Greenlee, Donald R. Commercial	SAIC One Skyline Tower, Suite 2200 5107 Leesburg Pike Falls Church, VA 22041 (703) 824-5909
Grussmeyer, Jimmy R. LtCol, US Army	ODUSA (OR) Pentagon, Room 2E660 Washington, DC 20310-0102 (703) 697-0366
Gunkel, Richard A. Commercial	Sigma Plus, Inc. 4001 Williamsburg Court Fairfax, VA 22032 (703) 591-0733
Gustafson, Ronald A. USAF Civilian	HQ AFOTEC/CNP P.O. Box 413 Kirtland AFB, NM 87117-7001 (505) 846-1844
Hall, David H. US Navy Civilian	Naval Weapons Center Code 3181 China Lake, CA 93555-6001 (619) 939-3383
Hamilton-Amthor,Jacqueline Commercial	International Test and Evaluation Assn 4400 Fairlakes Court Fairfax, VA 22033 (703) 356-4667
Hauze, Gene E. FFRDC	The MITRE Corporation 7525 Colshire Drive MS W958 McLean, VA 22102 (703) 883-7584
Hays, Johnny R. Maj, USAF	HQ SAC/XPSW 901 SAC Blvd, STE 2B5 Offutt AFB, NE 68113-5330 (402) 294-4220
Helmuth, Richard E. Commercial	Douglas Aircraft Company Mail Code 35-95 3855 Lakewood Blvd Long Beach, CA 90846 (213) 593-7241
Hinkle, Robert G. US Army Civilian	Office, Deputy Under Secy of the Army (OR) ATTN: SAUS-OR Pentagon, Room 1E643 Washington, DC 20310-0102 (703) 697-0367

Hires, Joyce A. US Army Civilian	US Army Test and Evaluation Command ATTN: AMSTE-TA-H Aberdeen Proving Ground, MD 21005-5055 (301) 278-5221
Hollis, Walter W. US Army Civilian	DUSA (OR), HQ Dept of the Army ATTN: SAUS(OR) Pentagon, Room 2E660 Washington, DC 20310-0102 (703) 695-0083
Jenkins, K. D. Capt, USMC	H&S Bn, MCOTEA 2080 Barksdale Place Quantico, VA 22134-5010 (703) 640-3141
Johnson, W. Donald US Army Civilian	US Army Materiel Systems Analysis Activity ATTN: AMXSY-J Aberdeen Proving Ground, MD 21005-5071 (301) 278-3701
Johnson, Roger W. Commercial	Institute for Simulation and Training 12424 Research Parkway, Suite 300 Orlando, FL 32826 (407) 658-5501
Jones, Raymond A. Commercial	Rockwell International Tactical Systems Advanced Development Engineering 1800 Satellite Blvd Duluth, GA 30136 (404) 497-5257
Kapos, Ervin Commercial	Kapos Associates Suite 1900 1101 Wilson Blvd Arlington, VA 22209-1603 703) 528-4575
Keck, Eric L. USAF Civilian	AFOTEC/OAN Kirtland AFB, NM 87117-7001 (505) 846-1938
Keegan, John H. US Navy Civilian	NUSC Code 38, Bldg 103 Newport, RI 02841-5047 (401) 841-3504
Kelley, John A., Jr. Commercial	Raytheon Missile Systems Lab (T3TA3) 50 Apple Mill Drive Tewksbury, MA 01876 (508) 858-9228
Kim, Dong S. US Army Civilian	US Army Concepts and Analysis Agency ATTN: CSCA-RST 8120 Woodmont Avenue Bethesda, MD 20814-2797 (301) 295-1652

King, Benjamin F.
US Army Civilian

US Army Materiel Systems Analysis Activity
ATTN : AMXSY-GS
Aberdeen Proving Ground, MD 21005-5071
(301) 278-4528

Kissin, Sidney
Other DoD

National Security Agency
ATTN: R55
9800 Savage Road
Fort George G. Meade, MD 20755-6000
(301) 688-0562

Knight, Wayne R.
US Army Civilian

US Army Operational Test & Evaluation Command
4501 Ford Avenue
Alexandria, VA 22302-1458
(703) 756-2446

Kocher, Theresa R.
US Army Civilian

HQ US Army Test and Evaluation Command
ATTN: AMSTE-TA-G
Aberdeen Proving Ground, MD 21005-5055
(301) 278-5221

Kolding, James C.
Maj, US Army

US Army TRADOC Analysis Center-WSMR
ATTN: ATRC-ZD
Bldg 1400
White Sands Missile Range, NM 88002-5502
(505) 678-7690

Kraft, Edward M.
Commercial

Calspan Corp/AEDC Div
MS-600
Arnold AFB, TN 37389-5000
(615) 454-6658

Kumashiro, Thomas Y
Col, USAF

HQ AFOTEC/TES
Kirtland AFB, NM 87117-7001
(505) 846-5211

Landweer, Philip R.
Commercial

BDM International
MS C-12
1801 Randolph Road, SE
Albuquerque, NM 87106
505) 848-5251

Lang, Paul M.
Commercial

SRS Technologies
1500 Wilson Blvd
Suite 800
Arlington, VA 22209
(703) 522-5588

Leach, James E.
Commercial

Lockheed-Sanders
MER 24 - 1282
P.O. Box 686
Nashua, NH 03061-0868
(603) 885-0176

Leahy, Michael J.
Maj, USMC

MCOTEA
Qtrs 4418A
Quantico, VA 22134-5017
(703) 640-3141

Lebowitz, Robert K. Commercial	MAR, Inc Suite 200 1800 SW 34th Street Fort Lauderdale, FL 33315 (305) 359-8150
Ledesma, Richard R. Other DoD	OUSD(A), DDDRE(T&E) Pentagon, Room 3E1060 Washington, DC 20301-3110 (703) 695-7171
Lee, Warren M. USAF Civilian	HQ SAC/XPSS 636 Forest Drive Offutt AFB, NE 68113-5001 (402) 294-4342
Lubas, Ann M. USMC Civilian	MCRDAC Program Analysis and Evaluation Code PAE Quantico, VA 22134 (703) 640-2427
MacFadzean, Robert H. US Navy Civilian	Naval Surface Warfare Center Code N-33 Dahlgren, VA 22448 (703) 663-1645
Marshall, John A. US Navy Civilian	Naval Air Test Center SA-UAV Patuxent River, MD 20670-5304 (301) 863-3922
McGill, Gregory K. USAF Civilian	HQ AFOTEC/ST 12510 Elena Drive, NE Kirtland AFB, NM 87117-7001 (505) 846-0631
Merryman, L. James LtCol, US Army	Joint Staff, J-8 Pentagon, Room 1D964 Washington, DC 20318-8000 (703) 614-3681
Merz, Vincent P. Commercial	VEDA, Inc. PAXR Division 300 Exploration Lexington Park, MD 20653 (301) 862-2100
Messing, Phil FFRDC	The Mitre Corporation 7525 Colshire Drive MS W952 McLean, VA 22102 (703) 883-5606
Metz, Thomas R Other DoD	OUSD(A)DDDRE(T&E/TFR) Pentagon, Room 3D1067 Washington, DC 20301-3110 (703) 697-4819

Metzger, James J. US Army Civilian	ODCSOPS, HQ Dept of the Army ATTN: DAMO-ZDS Washington, DC 20310-0410 (703) 695-2451
Meyer, Robert J. US Navy Civilian	Naval Weapons Center Code 3186 China Lake, CA 93555-6001 (619) 939-6686
Miller, Richard S. FFRDC	Institute for Defense Analyses 1801 N. Beauregard Street Alexandria, VA 22311 (703) 845-2264
Mixon, Donald E. FFRDC	The MITRE Corporation 7525 Colshire Drive McLean, VA 22102 (703) 883-6599
Mortin, David E. US Army Civilian	US Army Materiel Systems Analysis Activity ATTN: AMXSY-LX Aberdeen Proving Ground, MD 21005-5071 (301) 278-8786
Mullis, Charles W. US Army Civilian	STEWS-NR-PR P.O. Box 454 White Sands Missile Range, NM 88002 (505) 678-1221
Murashige, Allen M. USAF Civilian	HQ USAF/SAG Scientific & Technical Advisor Pentagon, Rm 1E389 Washington, DC 20330-5420 (703) 697-9824
Nichols, Beverly US Army Civilian	US Army Air Defense Artillery School Studies & Analysis, Dir Combat Developments ATTN: ATSA CDC S Fort Bliss, TX 79916-7050 (915) 568-6338
Niemeyer, William A. US Army Civilian	US Army Materiel Systems Analysis Activity ATTN: AMXSY-C Aberdeen Proving Ground, MD 21005-5071 (301) 278-6450
Nolan, Thomas E. US Army Civilian	US Army Materiel Systems Analysis Activity ATTN: AMXSY-CC Aberdeen Proving Ground, MD 21005-5071 (301) 278-6610
O'Neill, Patrick J. US Army Civilian	US Army Materiel Systems Analysis Activity ATTN: AMXSY-CA Aberdeen Proving Ground, MD 21005-6429 (301) 278-6429

Oppenheimer, David A. Commercial	SRA Corporation 2000 15th Street North Arlington, VA 22201 (703) 558-7694
Pace, Richard W. US Navy Civilian	Naval Air Test Center Range Directorate, RD(80) Bldg 1492 Patuxent River, MD 20670-1541 (301) 863-1550
Pacheco, Nelson S. FFRDC	Institute for Defense Analyses 1801 N. Beauregard Street Alexandria, VA 2231 (703) 845-2356
Paquette, Dale L. US Navy Civilian	Naval Underwater Systems Center Code 38202, Bldg 104 Newport, RI 02841-5047 (401) 841-4802
Parry, Thomas J. Other DoD	OASD(P&L)/L WSIO Pentagon, Room 2B322 Washington, DC 20301-1800 (703) 687-1366
Paulc, Eugene P. Capt, US Army	TRADOC Analysis Command - MTRY P.O. Box 8692 Monterey, CA 93943-0692 (408) 646-3086
Pfledderer, Chris USAF Civilian	ASD/ENSA Eglin AFB, FL 32542-5000
Plishker, R. Alan Commercial	International Test and Evaluation Assn 4400 Fair Lakes Court Fairfax, VA 22033-3899 (703) 631-6220
Polk, Sidney FFRDC	The MITRE Corporation MS Z509 7525 Colshire Drive McLean, VA 22102 (703) 883-6800
Prichard, David W. US Army Civilian	US Army Test and Evaluation Command ATTN: AMSTE-TA-W Aberdeen Proving Ground, MD 21005-5055 (301) 278-5278
Proctor, Michael D. LtCol, US Army	US Army TRADOC Analysis Command - MTRY P.O. Box 8692 Naval Post Graduate School Monterey, CA 93943-0692 (408) 646-3086

Reece, Rick L. Maj, USMC	H&S Bn, MCOTEA 8009 Pembroke Circle Quantico, VA 22134 (703) 640-3286
Reed, Stalker E. LtCol, USAF	AFOTEC/TEZ 6424 Colleen, NE Kirtland AFB, NM 87117-7001 (505) 846-5334
Reeves, John M. Commercial	SAIC Suite F 2301 Yale Blvd, SE Albuquerque, NM 87123 (505) 242-2828
Reig, Raymond W. US Army Civilian	Defense Systems Management College FD-TE Fort Belvoir, VA 22060-5426 (703) 664-5060
Reveal, Charles A. USAF Civilian	6545th Test Group/ENTL Dept of the Air Force Hill AFB, UT 84056 (801) 777-3824
Reynolds, Matthew T. US Navy Civilian	Naval Sea Systems Command Director, Test and Evaluation Office Washington, DC 20362 (703) 746-3020
Ritchie, Adelia E. Commercial	SAIC Suite 608 8619 Westwood Center Drive Vienna, VA 22182 (703) 749-5604
Rogers, Bruce F. Commercial	SAIC 8919 Westwood Center Drive Suite 608 Vienna, VA 22182 (703) 749-5610
Rosenberger, Jerome C. FFRDC	Institute for Defense Analyses 1801 N. Beauregard Street Alexandria, VA 22311 (703) 845-2095
Sanders, Patricia A. Other DoD	OASD (PA&E) Pentagon, Room 2B256 Washington, DC 20301-2700 (703) 697-3521
Schroffel, Steve A. Other Govt	US Immigration and Naturalization Service ATTN: CORAD, Room 6108 425 Eye Street, NW Washington, DC 20536 (202) 514-3343

Seglie, Ernest A. Other DoD	OSD (DOTE) Pentagon, Room 3E318 Washington, DC 20301-1700 (703) 697-7247
Severson, Tom US Navy Civilian	Naval Weapons Center Code 204 China Lake, CA 93555 (619) 939-3833
Simnacher, Gerald LtCol, USAF	HQ AFOTEC/ST 2070 A Mercury Drive, SE Kirtland AFB, NM 87117-7001 (505) 846-2585
Smith, Bruce L. LtCol, USAF	AFCSA/SAZ Pentagon, Room 1E386 Washington, DC 20330-5420 (703) 697-9827
Sorrells, Russell B. USAF Civilian	USAF/AFDC/DOF Route 2 238 Happy Valley Road Arnold AFB, TN 37389-5000 (615) 454-7721
Southard, Lounell D. US Army Civilian	US Army TRADOC Analysis Command-WSMR ATTN: ATRC-WBB White Sands Missile Range, NM 88002-5502 (505) 678-1461
Staniec, Cyrus J. Maj, US Army	OASD(PA&E) Land Forces Division Pentagon, Room 2B256 Washington, DC 20301-3147 (703) 697-1600
Starr, Stuart H. FFRDC	The MITRE Corporation 3422 Rustic Way Lane 7527 Colshire Drive McLean, VA 22102 (703) 883-5494
Stauch, Edward G. US Army Civilian	US Army Test and Evaluation Command ATTN: AMSTE-TA-S Aberdeen Proving Ground, MD 21005-5055 (301) 278-2775
Thomas, Clayton J. USAF Civilian	HQ USAF/AFSAA Pentagon, Room 1E386 Washington, DC 20330-5420 (703) 697-4300
Tino, Marshall J. US Navy Civilian	Naval Surface Warfare Center Protection System Dept Code H Dahlgren, VA 22448-5000 (703) 663-8771

Toohill, Barbara G. FFRDC	The MITRE Corporation MS W410 7525 Colshire Drive McLean, VA 22102 (703) 883-6454
Toole, Michael T. Col, US Army	Strategic Defense Initiative Org/TNE Pentagon, Room 1E180-C Washington, DC 20301-7100 (703) 693-1577
Tranby, Erling D. US Navy Civilian	Naval Ship Weapons Systems Engineer Sta Code 4L00 Port Hueneme, CA 93043-5007 (805) 982-0001
Tucker, Dewey D. LtCol, USMC	OASD (PA&E) Pentagon, Room 2B256 Washington, DC 20301-1800 (703) 697-7085
Tucker, Jesse R. LtCol, USAF	Defense Information Systems Agency 3701 North Fairfax Drive Arlington, VA 22203 (703) 696-1841
Turner, Stewart W. Commercial	WNG Associates, Inc. P.O. Box 1382 Fort Walton Beach, FL 32549-0719 (904) 243-1313
Veit, Clairice T. FFRDC	The RAND Corporation 1700 Main Street Santa Monica, CA 90401 (213) 393-0411
Victor, Alfred E. FFRDC	Institute for Defense Analyses 2001 N. Beauregard Street Alexandria, VA 22311-1772 (703) 845-2481
Walker, John K., Jr. Commercial	Military Operations Research Society 101 S. Whiting Street Suite 202 Alexandria, VA 22304 (703) 751-7290
Waller, Sylvia L. FFRDC	Institute for Defense Analyses 2001 N. Beauregard Street Alexandria, VA 22311-1772 (703) 845-2344
Warner, Steve FFRDC	Institute for Defense Analyses 2001 N. Beauregard Street Alexandria, VA 22311-1772 (703) 845-2096

West, Larry L. US Army Civilian	US Army Test and Evaluation Command ATTN: AMSTE-TA-G Aberdeen Proving Ground, MD 21005-5055 (301) 278-2456
White, Allan E. Commercial	Kapos Associates Suite 1900 1101 Wilson Blvd Arlington, VA 22209 (703) 528-4575
Wiles, Richard I. Commercial	Military Operations Research Society 101 S Whiting Street Suite 202 Alexandria, VA 22304 (703) 751-7290
Williams, Marion L. USAF Civilian	HQ AFOTEC Chief Scientist Kirtland AFB, NM 87117-7001 (505) 846-0607
Williams, Stephen D. LtCol, US Army	US Army War College ATTN: AWCAW Bldg 257 Carlisle, PA 17013-5050 (717) 245-4169
Williams, Douglas P. Capt, USMC	Department of the Navy USMC (DRPM AAA) Direct Reporting Program Manager Washington, DC 20380-0001 (703) 696-1116
Wisnowski, James W. Capt, USAF	HQ AFOTEC/LG4W 11816 Caribou Ave, NE Kirtland AFB, NM 87117-7001 (505) 846-5644
Wood, Ronald L. Commercial	Research Analysis & Maintenance, Inc. Vista Hills Bank Plaza 1790 Lee Trevino Drive, Suite 600 El Paso, TX 79936 (915) 592-7047
Worlein, Cheryl K. Lt, US Navy	Operational Test & Evaluation Force NAS, North Island San Diego, CA 92135 (619) 545-2724
Youngblood, James E. US Army Civilian	US Army Materiel Systems Analysis Acty ATTN: AMXSY-ADH Aberdeen Proving Ground, MD 21005-5071 (301) 278-6373
Zang, Charles W. Commercial	BDM International, Inc. 1501 BDM Way McLean, VA 22102 (703) 848-5249